The Astrophysical Journal, 595:531–549, 2003 September 20

THE SECULAR EVOLUTION OF THE PRIMORDIAL KUIPER BELT

Joseph M. Hahn[1]
Lunar and Planetary Institute, 3600 Bay Area Boulevard, Houston, TX 77058; hahn@lpi.usra.edu
Received 2003 April 5; accepted 2003 May 29

ABSTRACT

A model that rapidly computes the secular evolution of a gravitating disk-planet system is developed. The disk is treated as a nested set of gravitating rings, with the rings'/planets' time evolution being governed by the classical Laplace-Lagrange solution for secular evolution but modified to account for the disk's finite thickness h. The Lagrange planetary equations for this system yield a particular class of spiral wave solutions, usually called apsidal density waves and nodal bending waves. There are two varieties of apsidal waves—long waves and short waves. Planets typically launch long density waves at the disk's nearer edge or else at a secular resonance in the disk, and these waves ultimately reflect downstream at a more distant disk edge or else at a Q barrier in the disk, whereupon they return as short density waves. Planets also launch nodal bending waves, and these have the interesting property that they can stall in the disk, that is, their group velocity plummets to zero upon approaching a region in the disk that is too thick to support further propagation of bending waves. The rings model is used to compute the secular evolution of a Kuiper Belt having a variety of masses, and it is shown that the early massive belt was very susceptible to the propagation of low-amplitude apsidal and nodal waves launched by the giant planets. For instance, these waves typically excited orbits to $e \sim \sin i \sim 0.01$ in a primordial Kuiper Belt of mass $M_{\rm KB} \sim 30$ Earth masses. Although these orbital disturbances are quite small, the resulting fractional variations in the disk's surface density due to the short density waves is usually large, typically of order unity. This epoch of apsidal and nodal wave propagation probably lasted throughout the Kuiper Belt's first $\sim 10^7$ to $\sim 5 \times 10^8$ yr, with the waves being shut off between the time when the large $R \gtrsim 100$ km Kuiper Belt objects first formed and when the belt was subsequently eroded and stirred up to its present configuration.

Subject headings: celestial mechanics — Kuiper Belt — planetary systems: protoplanetary disks — solar system: formation

On-line material: mpg animation

1. INTRODUCTION

The Kuiper Belt is a vast swarm of comets orbiting at the solar system's outer edge. This belt is composed of debris that was left over from the epoch of planet formation, and this swarm's distribution of orbit elements preserves a record of events that occurred when the solar system was still quite young. A common goal of nearly all dynamical studies of the Kuiper Belt is to decipher this record. However, the record is still open to some interpretation.

The dots in Figure 1 show the Kuiper Belt object (KBO) eccentricities e and inclinations i versus their semimajor axes a. This figure reveals the KBOs' three major dynamical classes: the Plutinos, which inhabit Neptune's 3 : 2 resonance at $a = 39.5$ AU; the "Main Belt" KBOs, which are the nonresonant KBOs orbiting between 40 AU $\lesssim a \lesssim$ 48 AU; and the more distant "Scattered Belt" KBOs, which live in eccentric, nearly Neptune-crossing orbits. The figure also shows that the Plutinos and the Scattered KBOs have inclinations that span $0^\circ \lesssim i \lesssim 30^\circ$, while the Main Belt KBOs appear to have a bimodal distribution of inclinations centered on $i \simeq 2^\circ$ and $i \simeq 17^\circ$ (Brown 2001). Note that accretion models show that these large $\sim$100+ km KBOs must have first formed from much smaller planetesimal seeds that were initially in nearly circular and coplanar orbits having e and $\sin i \lesssim 0.001$ (Kenyon & Luu 1999). However, gravitational self-stirring cannot account for the Kuiper Belt's current excited state, so one or more mechanisms must also have stirred up the Kuiper Belt since the time of formation.

The orbits of the Scattered KBOs are perhaps the most easily understood. These objects have likely had one or more close encounters with Neptune, which lofted these bodies into eccentric, inclined orbits (Duncan & Levison 1997). Repeated encounters with Neptune cause these objects' semimajor axes and eccentricities to evolve stochastically along the Neptune-crossing curve shown in Figure 1, and most of these bodies are ultimately ejected or accreted by the giant planets. However, Neptune has numerous weak, high-order mean motion resonances that thread the Kuiper Belt, and these resonances permit some of these Scattered Belt objects to diffuse to lower eccentricities. This allows a small percentage of the Scattered Belt objects to persist over the age of the solar system at eccentricities just below the Neptune-crossing curve seen in Figure 1 (Duncan & Levison 1997). This diffusion to lower eccentricities might also have been more pronounced if Neptune's orbit also migrated outward. In particular, Gomes (2003) shows that during the epoch of planet migration, mean motion and secular resonances act as pathways that allow some scattered KBOs to descend *irreversibly* into lower eccentricity orbits that are far from the Neptune-crossing curve and hence stable. Since Scattered Belt KBOs have large inclinations of $i \gtrsim 10^\circ$, this process might also account for the Main Belt's bimodal inclination distribution, with the $i \sim 2^\circ$ component

[1] Current address: Institute for Computational Astrophysics, Department of Astronomy and Physics, Saint Mary's University, Halifax, NS B3H 3C3, Canada; jhahn@olympus.stmarys.ca.

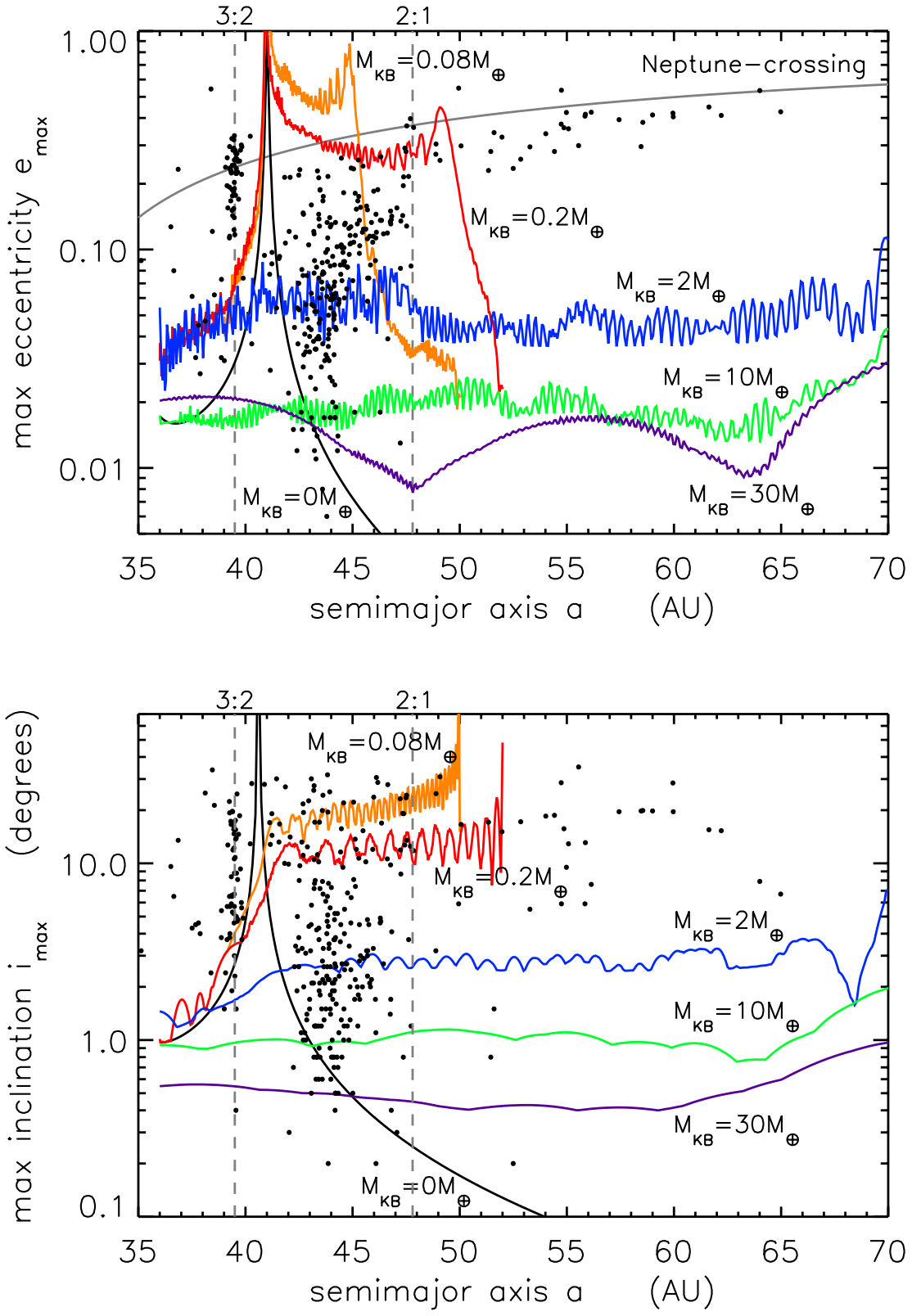

FIG. 1.—Maximum eccentricities, e_{max}, and inclinations, i_{max}, vs. semimajor axis a for simulations having a variety of Kuiper Belt masses, M_{KB}; see § 4.2 for model details. Dots indicate the orbits of 340 KBOs observed over multiple oppositions, with orbits provided by the Minor Planet Center. The locations of the 3 : 2 and 2 : 1 resonances are indicated, and orbits above the gray curve are Neptune crossing.

representing the Main Belt's native population of low-inclination KBOs, and the $i \sim 17°$ being due to scattered invaders who were deposited in the Main Belt by Neptune. However, this process is quite inefficient, since only $\varepsilon \sim 0.1\%$ of these Scattered KBOs manage to find orbits that are stable over a solar age (Gomes 2003).

The possibility that Neptune's orbit has expanded outward is also supported by the cluster of KBOs that inhabit

Neptune's 3 : 2 resonance (see Fig. 1). This may have occurred when Neptune first formed and began to vigorously scatter the local planetesimal debris. This process can drive an exchange of angular momentum between the planets and the planetesimal disk (Fernandez & Ip 1984), so this episode of disk clearing could have resulted in a rearrangement of all of the giant planets' orbits over a timescale of $\sim 10^7$ yr (Hahn & Malhotra 1999). Outward planet migration would also cause Neptune's mean-motion resonances to sweep out across the primordial Kuiper Belt, and these migrating resonances are quite effective at capturing KBOs and pumping up their eccentricities (Malhotra 1995). Models of this process show that if Neptune's orbit had in fact smoothly expanded some $\Delta a \sim 7$ AU over a timescale longer than a few million years, then resonance capture would have deposited numerous KBOs in the 3 : 2 and 2 : 1 resonances with eccentricities distributed over $0 \lesssim e \lesssim 0.3$ (Malhotra 1995; Chiang & Jordan 2002). Although numerous KBOs do indeed inhabit Neptune's 3 : 2 resonance, only a handful are known to live near the 2 : 1 resonance at $a = 47.8$ AU, and many of these bodies have eccentricities of $e \sim 0.3$, which puts them quite near the Neptune-crossing curve (see Fig. 1). Thus, it is possible that some or all of the bodies orbiting near $a = 47.8$ AU might instead be members of the Scattered Belt. If planet migration did indeed occur, then the apparent low abundance of KBOs at the 2 : 1 resonance is a mystery that can only be partly due to the observational bias that selects against the discovery of lower eccentricity objects at the 2 : 1 resonance (Jewitt, Luu, & Trujillo 1998).

Of particular interest here is the Main Belt, which preserves additional evidence for another mechanism having stirred up the Kuiper Belt. Again, if Neptune did indeed migrate outward $\Delta a \sim 7$ AU, then the entire Main Belt was swept by the advancing 2 : 1 resonance. Models of planet migration show that the efficiency of resonance capture is no more than $\sim$50% (Chiang & Jordan 2002), so the current members of the Main Belt evidently avoided permanent capture by slipping through the advancing 2 : 1 resonance. However, this planet migration scenario is utterly unable to account for the high inclinations of $0^\circ \lesssim i \lesssim 30^\circ$ observed in the Main Belt (see Fig. 1), since the N-body simulations show that the advancing 2 : 1 resonance typically excites Main Belt inclinations by only a few degrees (Malhotra 1995; Chiang & Jordan 2002). Evidently, an additional mechanism is also responsible for exciting the Main KBO Belt. It has been shown that sizable KBO excitation could occur if a recently formed Neptune had been scattered outward by Jupiter and/or Saturn into a belt-crossing orbit (Thommes, Duncan, & Levison 2002). Another possible source of KBO excitation is the invasion of the Main Belt by the Scattered Belt (Gomes 2003). However, this latter process is a very inefficient mechanism, having $\varepsilon \sim 0.001$; an alternate mechanism that is possibly more efficient is explored below.

It has also been suggested that secular resonance sweeping may have been responsible for exciting the Kuiper Belt (Nagasawa & Ida 2000). Note that the locations of the secular resonances are very sensitive to the solar system's mass distribution. Consequently, the depletion of the solar nebula (which includes perhaps $\sim$99% of the solar system's initial mass content) could have driven these secular resonances across vast tracts of the solar system. Indeed, the model by Nagasawa & Ida (2000) suggests that if the nebula was depleted on a timescale of $\tau \sim 10^7$ yr, Main Belt inclinations of $i \sim 20^\circ$ could have been excited by the passage of the ν_{15} secular resonance as it swept outward to infinity as the nebula was depleted.[2] However, this model is rather idealized in that it treats the Kuiper Belt as massless, which is a concern since a primordial Kuiper Belt having some mass is also susceptible to the propagation of very long wavelength spiral waves that could be launched at a secular resonance in the disk (Ward & Hahn 1998a, 2003). This issue is worth further examination, since wave action can alter the magnitude of resonant excitation considerably (Hahn, Ward, & Rettig 1995; Ward & Hahn 1998a). But even if there is no resonance in the disk, planets orbiting interior to a particle disk can still launch these spiral waves at the disk's inner edge (Ward & Hahn 1998b). Preliminary results from a model of secular resonance sweeping also reveals that a Kuiper Belt having only a modest amount of mass is utterly awash in these waves once the nebula is depleted (Hahn & Ward 2002). However, the purpose of the present study is first to characterize the properties of these waves in the simpler, postnebula environment, and then to explore their cosmogonic implications. Thus, the following will consider a suite of models of the secular evolution of the outer solar system for primordial Kuiper Belts having a variety of masses.

Accretion models tell us that the primordial Kuiper Belt must have had a mass of $M_{\rm KB} \sim 30\ M_\oplus$ in the 30 AU $< a <$ 50 AU interval in order for Pluto and its cohort of KBOs to have formed and survived over the age of the solar system (Kenyon & Luu 1999). A similar Kuiper Belt mass is also needed to drive Neptune's orbital migration of $\Delta a \sim 7$ AU (Hahn & Malhotra 1999). However, the current mass is $M_{\rm KB} \sim 0.2\ M_\oplus$ (Jewitt, Luu, & Trujillo 1998), so the Kuiper Belt appears to have been eroded by a factor of $\sim$150. This may be due to a dynamical erosion of the belt by Neptune or possibly by other perturbers that may once have been roaming about the outer solar system, as well as due to the collisional erosion that has since ground all of the smaller KBOs down to dust grains that are then removed from the solar system by radiation forces (Kenyon & Luu 1999; Kenyon & Bromley 2001). The following will give results obtained from models of the secular evolution of the outer solar system for Kuiper Belts having masses in the interval $0 \leq M_{\rm KB} \leq 30\ M_\oplus$.

Section 2 derives the so-called rings model that will be used to study the secular evolution of disk-planet systems; the reader uninterested in these details might skip ahead to § 3 or § 4. Since spiral density and bending waves appear prominently in the model results, their properties are examined in § 3. Section 4 describes the model's application to the primordial Kuiper Belt, and a summary of results is then given in § 5.

[2] It should be noted that the resulting inclination excitation is also sensitive to the tilt between the nebula midplane and the invariable plane. For instance Nagasawa & Ida (2000) place the nebula midplane in the ecliptic, and this results in substantial excitation. But if the nebula midplane is instead placed in the invariable plane, which is tilted 1°.6 from the ecliptic, then almost no excess excitation results (Hahn & Ward 2002). We also note that the nebula models of Nagasawa & Ida (2000) as well as Hahn & Ward (2002) both treat the gas disk as a rigid slab of gas. However, a more realistic treatment would allow the nebula disk to flex and warp in response to the planets' secular perturbations. It is suspected that this additional degree of freedom will substantially alter the secular resonance sweeping; indeed, it can be argued that the ν_{15} never did sweep across the Kuiper Belt on account of this flexure (E. Chiang & W. Ward 2002, private communication), so perhaps secular resonance sweeping of the Kuiper Belt is actually a moot issue.

2. THE SECULAR EVOLUTION OF DISK-PLANET SYSTEMS

This section treats the disk as a collection of nested gravitating rings in orbit about the Sun. Their mutual perturbations will cause these rings to slowly flex and tilt over time, and this evolution is governed by the Lagrange planetary equations.

2.1. *The Rings Model*

Begin with the gravitational potential that a single perturbing ring of mass m' exerts at the point $\boldsymbol{r}$ on another ring of mass m:

$$\Phi'(\boldsymbol{r}) = -\int \frac{G\rho'\,dV'}{\Delta}\ , \tag{1}$$

where G is the gravitation constant, ρ' is the mass density of the differential volume element dV', Δ is the separation between the perturbing mass element $\rho' dV'$ at $\boldsymbol{r}'$ and the field point $\boldsymbol{r}$, and the integration proceeds over the three-dimensional extent of ring m'. Hereafter, primed quantities refer to the perturbing ring m', and unprimed quantities refer to the perturbed ring m. Each ring can be thought of as a swarm of numerous particles all having a common semimajor axis a and an identical mean orbital eccentricity e, inclination i, longitude of periapse $\tilde{\omega}$, and longitude of ascending node Ω. It is also assumed that these particles have an isotropic dispersion velocity c that gives rise to the ring's finite radial and vertical half-thickness $h \simeq c/n$, where n is the ring's mean motion. It is also assumed that the density ρ' varies only in the azimuthal direction because of the Keplerian motion of the ring's particles; in this case the density is $\rho' = \lambda'/4h'^2$, where

$$\lambda' = \frac{m'r'}{2\pi a'^2\sqrt{1-e'^2}} \tag{2}$$

is the ring's azimuthal mass per unit length (Murray & Dermott 1999). In cylindrical coordinates $\boldsymbol{r}' = (l', \phi', z')$ and $dV' = l'\,dl'\,d\phi'\,dz'$. If the slight radial variations in the integrand of equation (1) are ignored (i.e., $\int l'\,dl' \simeq 2r'h'$), the potential becomes

$$\Phi'(\boldsymbol{r}) \simeq -\int_{-\pi}^{\pi} d\phi' \int_{z_0'-h'}^{z_0'+h'} dz' \frac{G\lambda' r'}{2h'\Delta}\ , \tag{3}$$

where $z_0'(\phi')$ is the longitude-dependent height of the perturbing ring's midplane from the $z = 0$ plane. Of course, the perturbed ring m also has a radial and vertical half-thickness h, and it is useful to form an effective potential by averaging $\Phi'(\boldsymbol{r})$ over the radial and vertical extent of ring m:

$$\langle\Phi'(\boldsymbol{r})\rangle = \int_{-h}^{h}\frac{dl}{2h}\int_{z_0-h}^{z_0+h}\frac{dz}{2h}\Phi'(\boldsymbol{r}) \equiv -\int_{-\pi}^{\pi} d\phi' \frac{G\lambda' r'}{r} Q\ , \tag{4}$$

where

$$Q \simeq \int_{z_0-h}^{z_0+h}\frac{dz}{2h}\int_{z_0'-h'}^{z_0'+h'}\frac{dz'}{2h'}\frac{r}{\Delta}\ , \tag{5}$$

where again the slight variations in the integrand with l are ignored, and only Δ is assumed to be sensitive to the variations in z and z'. Note that this averaging of Φ' is essential in order for the algorithm developed below to conserve angular momentum.

The next task is to evaluate the double integrals in Q. First note that the separation Δ between the perturbing mass element $\rho'\,dV'$ at $\boldsymbol{r}'$ and the field point $\boldsymbol{r}$ obeys $\Delta^2 = r^2 + r'^2 - 2[(r^2 - z^2)(r'^2 - z'^2)]^{1/2}\cos(\phi' - \phi) - 2zz'$, where $r^2 = l^2 + z^2$. Setting $\alpha \equiv r'/r$, $\beta \equiv z/r$, and $\beta' \equiv z'/r'$, then

$$\left(\frac{\Delta}{r}\right)^2 \simeq 1 + \alpha^2 - 2\alpha\cos(\phi' - \phi) + (\beta^2 + \beta'^2)\alpha\cos(\phi' - \phi) - 2\alpha\beta\beta' \tag{6}$$

to second order in β, which are of order the rings' inclinations and are assumed to be small. Inserting this into the expression for Q yields

$$\begin{aligned} Q &= \frac{1}{4\mathfrak{h}\mathfrak{h}'}\int_{\beta_0-\mathfrak{h}}^{\beta_0+\mathfrak{h}} d\beta \int_{\beta_0'-\mathfrak{h}'}^{\beta_0'+\mathfrak{h}'} d\beta' \Big[1 + \alpha^2 - 2\alpha\cos\Delta\phi \\ &\quad + (\beta^2 + \beta'^2)\alpha\cos\Delta\phi - 2\alpha\beta\beta'\Big]^{-1/2} \\ &= \int_{\beta_0'-\mathfrak{h}'}^{\beta_0'+\mathfrak{h}'} \frac{\ln(\Upsilon)d\beta'}{4\mathfrak{h}\mathfrak{h}'\sqrt{\alpha\cos\Delta\phi}}\ , \end{aligned} \tag{7}$$

where $\mathfrak{h} \equiv h/r \simeq h/a$ and $\mathfrak{h}' \equiv h'/r' \simeq h'/a'$ are the fractional half-thicknesses of rings m and m', $\beta_0 \equiv z_0/r$ and $\beta_0' \equiv z_0'/r'$ are the rings' midplane latitudes, $\Delta\phi \equiv \phi' - \phi$, and the right-hand side of equation (7) is the result of doing the integration in β, where

$$\Upsilon = \frac{\beta'\alpha - \beta_0\alpha\cos\Delta\phi - \mathfrak{h}\alpha\cos\Delta\phi - \sqrt{\alpha\cos\Delta\phi(\Gamma + \varepsilon)}}{\beta'\alpha - \beta_0\alpha\cos\Delta\phi + \mathfrak{h}\alpha\cos\Delta\phi - \sqrt{\alpha\cos\Delta\phi(\Gamma - \varepsilon)}}\ , \tag{8}$$

with $\Gamma = 1 + \alpha^2 - 2\alpha\cos\Delta\phi + (\beta_0^2 + \mathfrak{h}^2 + \beta'^2)\alpha\cos\Delta\phi - 2\alpha\beta'\beta_0$ and $\varepsilon = 2\mathfrak{h}(\beta_0\cos\Delta\phi - \beta')\alpha$. The β and the $\mathfrak{h}$ are assumed to be small, so ε is second order in the small quantities and is negligible when compared to other terms. Thus,

$$\begin{aligned} \Upsilon &\simeq \frac{1 - [\beta'\alpha - \beta_0\alpha\cos\Delta\phi - \mathfrak{h}\alpha\cos\Delta\phi]/\sqrt{\alpha\cos\Delta\phi\Gamma}}{1 - [\beta'\alpha - \beta_0\alpha\cos\Delta\phi + \mathfrak{h}\alpha\cos\Delta\phi]/\sqrt{\alpha\cos\Delta\phi\Gamma}} \\ &\simeq 1 + 2\mathfrak{h}\sqrt{\frac{\alpha\cos\Delta\phi}{\Gamma}}\ , \end{aligned} \tag{9}$$

so $\ln\Upsilon \simeq 2\mathfrak{h}(\alpha\cos\Delta\phi/\Gamma)^{1/2}$. This is inserted back into equation (7) and the remaining integral over β' is evaluated similarly, yielding $Q \simeq \ln(\Lambda)/2\mathfrak{h}'(\alpha\cos\Delta\phi)^{1/2}$, where

$$\begin{aligned} \Lambda &= \left[1 - \frac{\beta_0\alpha - \beta_0'\alpha\cos\Delta\phi - \mathfrak{h}'\alpha\cos\Delta\phi}{\sqrt{\alpha\cos\Delta\phi(\Psi + \mathscr{Z} + \xi)}}\right] \\ &\quad \times \left[1 - \frac{\beta_0\alpha - \beta_0'\alpha\cos\Delta\phi + \mathfrak{h}'\alpha\cos\Delta\phi}{\sqrt{\alpha\cos\Delta\phi(\Psi + \mathscr{Z} - \xi)}}\right]^{-1} \\ &\simeq 1 + 2\mathfrak{h}'\sqrt{\frac{\alpha\cos\Delta\phi}{\Psi + \mathscr{Z}}}\ , \end{aligned} \tag{10}$$

where $\Psi \equiv 1 + \alpha^2 - 2\alpha\cos\Delta\phi(1 - H^2) \simeq (1 + \alpha^2)(1 + H^2) - 2\alpha\cos\Delta\phi$, $H^2 \equiv (\mathfrak{h}^2 + \mathfrak{h}'^2)/2$, $\mathscr{Z} = (\beta_0^2 + \beta_0'^2)\alpha\cos\Delta\phi - 2\alpha\beta_0\beta_0'$, and $\xi \equiv 2\mathfrak{h}'(\beta_0'\cos\Delta\phi - \beta_0)\alpha$ is another negligible term. Inserting $\ln\Lambda \simeq 2\mathfrak{h}'[\alpha\cos\Delta\phi/(\Psi + \mathscr{Z})]^{1/2}$ back into Q yields

$$Q \simeq \frac{1}{\sqrt{\Psi + \mathscr{Z}}} \simeq \Psi^{-1/2} - \frac{1}{2}\mathscr{Z}\Psi^{-3/2}\ . \tag{11}$$

A Fourier expansion of Ψ^{-s} will be useful, i.e., $\Psi^{-s} = \frac{1}{2}\sum_{m=-\infty}^{\infty} \tilde{b}_s^{(m)} \cos m(\phi' - \phi)$, where

$$\tilde{b}_s^{(m)}(\alpha, \mathfrak{h}, \mathfrak{h}') = \frac{2}{\pi}\int_0^{\pi} \frac{\cos(m\phi)d\phi}{\{(1+\alpha^2)[1+\frac{1}{2}(\mathfrak{h}^2+\mathfrak{h}'^2)] - 2\alpha\cos\phi\}^s} \tag{12}$$

is the softened Laplace coefficient. The usual unsoftened form is $b_s^{(m)}(\alpha) = \tilde{b}_s^{(m)}(\alpha, 0, 0)$. These two coefficients are nearly equal when α is far from unity, but the softened form is finite at $\alpha = 1$, whereas the unsoftened form diverges. Writing Q in terms of softened Laplace coefficients thus gives

$$Q \simeq \tfrac{1}{2}\sum_{m=-\infty}^{\infty} \cos m(\phi' - \phi) \times \left\{\tilde{b}_{1/2}^{(m)} - \left[\tfrac{1}{2}(\beta_0^2 + \beta_0'^2)\alpha\cos(\phi' - \phi) - \beta_0\beta_0'\alpha\right]\tilde{b}_{3/2}^{(m)}\right\}, \tag{13}$$

which is then inserted in equation (4) to get the perturbing ring's potential,

$$\langle\Phi'\rangle = -\int_{-\pi}^{\pi} d\phi' \frac{G\lambda' r'}{2r}\sum_{m=-\infty}^{\infty} \cos m(\phi' - \phi) \times \left\{\tilde{b}_{1/2}^{(m)} - \left[\frac{1}{2}(\beta_0^2 + \beta_0'^2)\alpha\cos(\phi' - \phi) - \alpha\beta_0\beta_0'\right]\tilde{b}_{3/2}^{(m)}\right\}. \tag{14}$$

The final task of this section is to write the ring's coordinates in terms of its orbit elements,

$$r \simeq a\left(1 - e\cos\nu - \tfrac{1}{2}e^2 + \tfrac{1}{2}e^2\cos 2\nu\right), \tag{15a}$$

$$\phi = \tilde{\omega} + \nu \simeq \tilde{\omega} + M + 2e\sin M, \tag{15b}$$

$$\beta_0 = \frac{z_0}{r} \simeq \sin i \sin(\phi - \Omega), \tag{15c}$$

where ν is the true anomaly of a ring element at $\boldsymbol{r}$, and $M = nt$ is the corresponding mean anomaly. Inserting equations (15) into the potential $\langle\Phi'\rangle$, expanding to second order in e and i, doing the ϕ' integration in equation (14), and time averaging the resulting expression over the orbital period of ring m then yields the time-averaged potential $\langle\bar{\Phi}'\rangle$ experienced by ring m due to ring m' assuming small eccentricities and inclinations:

$$\langle\bar{\Phi}'\rangle = -\frac{Gm'}{a}\left[\frac{1}{2}\tilde{b}_{1/2}^{(0)} + \frac{1}{8}(e^2 + e'^2)f + \frac{1}{4}ee'\cos(\tilde{\omega}' - \tilde{\omega})g - \frac{1}{8}(i^2 + i'^2)\alpha\tilde{b}_{3/2}^{(1)} + \frac{1}{4}ii'\cos(\Omega' - \Omega)\alpha\tilde{b}_{3/2}^{(1)}\right], \tag{16}$$

where the f and g functions are

$$f(\alpha, \mathfrak{h}, \mathfrak{h}') = \left(2\alpha\frac{\partial}{\partial\alpha} + \alpha^2\frac{\partial^2}{\partial\alpha^2}\right)b_{1/2}^{(0)} = \alpha\tilde{b}_{3/2}^{(1)} - 3\alpha^2H^2(2 + H^2)\tilde{b}_{5/2}^{(0)}, \tag{17a}$$

$$g(\alpha, \mathfrak{h}, \mathfrak{h}') = \left(2 - 2\alpha\frac{\partial}{\partial\alpha} - \alpha^2\frac{\partial^2}{\partial\alpha^2}\right)b_{1/2}^{(1)} = -\alpha\tilde{b}_{3/2}^{(2)} + 3\alpha^2H^2(2 + H^2)\tilde{b}_{5/2}^{(1)}, \tag{17b}$$

where $H^2 = \frac{1}{2}(\mathfrak{h}^2 + \mathfrak{h}'^2)$ and α has been redefined as $\alpha = a'/a$. The right-hand side of equations (17) is derived in Appendix A, and it is shown in Appendix B that the softened Laplace coefficients can be written in terms of complete elliptic integrals. Consequently, functions f, g, and $\tilde{b}_s^{(m)}$ can all be rapidly evaluated without relying on a numerical integration of equation (12). Also note that the $\tilde{b}_s^{(m)}$, f, and g functions obey the following reciprocal relations:

$$\tilde{b}_s^{(m)}(\alpha^{-1}, \mathfrak{h}', \mathfrak{h}) = \alpha^{2s}\tilde{b}_s^{(m)}(\alpha, \mathfrak{h}, \mathfrak{h}'), \tag{18a}$$

$$f(\alpha^{-1}, \mathfrak{h}', \mathfrak{h}) = \alpha f(\alpha, \mathfrak{h}, \mathfrak{h}'), \tag{18b}$$

$$g(\alpha^{-1}, \mathfrak{h}', \mathfrak{h}) = \alpha g(\alpha, \mathfrak{h}, \mathfrak{h}'). \tag{18c}$$

These relations are used in § 2.2.1 to show that the equations of motion developed below conserve angular momentum.

The laborious procedure of expanding, integrating, and then time-averaging $\langle\Phi'\rangle$ is not included here, since a similar analysis can be found in Murray & Dermott (1999).[3] Since the terms proportional to e'^2 and i'^2, as well as the first term in equation (16), do not contribute to the resulting dynamical equations, they may be neglected. The disturbing function R for ring m due to another ring m' is $-1\times$ the surviving terms in $\langle\bar{\Phi}'\rangle$, i.e.,

$$R = \frac{Gm'}{a}\left[\frac{1}{8}fe^2 + \frac{1}{4}gee'\cos(\tilde{\omega}' - \tilde{\omega}) - \frac{1}{8}\alpha\tilde{b}_{3/2}^{(1)}i^2 + \frac{1}{4}\alpha\tilde{b}_{3/2}^{(1)}ii'\cos(\Omega' - \Omega)\right]. \tag{19}$$

Note that when $\mathfrak{h} = \mathfrak{h}' = 0$, the disturbing function for a point mass m perturbed by point mass m' is recovered (Brouwer & Clemence 1961; Murray & Dermott 1999), which is to be expected since both a point mass and a thin ring have the same disturbing function to this degree of accuracy (Murray & Dermott 1999).

It should also be noted that many celestial mechanics texts develop two distinct expressions for the disturbing function R, one due to a perturber in an interior orbit having $a' < a$, and another expression due to an exterior perturber having $a' > a$ (e.g., Brouwer & Clemence 1961; Murray & Dermott 1999). However, this pairwise development is unnecessary in this application since equation (19) is valid for $\alpha = a'/a < 1$ as well as for $\alpha > 1$. Indeed, it is straightforward to show that these pairs of disturbing functions, such as equations (7.6) and (7.7) in Murray & Dermott (1999), are in fact equivalent to equation (19) with $\mathfrak{h} = \mathfrak{h}' = 0$; they only appear distinct, since one is a function of α and the other is actually a function of α^{-1}.

In terms of the variables

$$h = e\sin\tilde{\omega}, \quad p = i\sin\Omega, \tag{20a}$$

$$k = e\cos\tilde{\omega}, \quad q = i\cos\Omega, \tag{20b}$$

the disturbing function can then be written

$$R = n^2a^2\left(\frac{m'}{M_\odot + m}\right)\left[\frac{1}{8}f(h^2 + k^2) + \frac{1}{4}g(hh' + kk') - \frac{1}{8}\alpha\tilde{b}_{3/2}^{(1)}(p^2 + q^2) + \frac{1}{4}\alpha\tilde{b}_{3/2}^{(1)}(pp' + qq')\right], \tag{21}$$

where the mean motion $n = [G(M_\odot + m)/a^3]^{1/2}$, with $M_\odot$ being the solar mass.

[3] Actually, Murray & Dermott (1999) derive the time-averaged acceleration (rather than a potential) that ring m' exerts on m, which they insert into the Gauss equations to obtain a set of dynamical equations equivalent to that obtained here when $\mathfrak{h} = 0$.

2.1.1. *The N-Ring Problem*

For the more general problem of N perturbing rings, replace the perturbing ring mass m' with m_k and give all the other primed quantities the subscript k. The disturbing function $R \rightarrow R_j$ for the perturbed ring having a mass $m \rightarrow m_j$ is sum of equation (21) over all the other rings $k \neq j$. Setting $\alpha_{jk} \equiv a_k/a_j$, $n_j = [G(M_\odot + m_j)/a_j^3]^{1/2}$, and

$$A_{jj} = \frac{n_j}{4} \sum_{k \neq j} \left(\frac{m_k}{M_\odot + m_j} \right) f(\alpha_{jk}, \mathfrak{h}_j, \mathfrak{h}_k) \ , \tag{22a}$$

$$A_{jk} = \frac{n_j}{4} \left(\frac{m_k}{M_\odot + m_j} \right) g(\alpha_{jk}, \mathfrak{h}_j, \mathfrak{h}_k), \quad j \neq k \ , \tag{22b}$$

$$B_{jj} = -\frac{n_j}{4} \sum_{k \neq j} \left(\frac{m_k}{M_\odot + m_j} \right) \alpha_{jk} \tilde{b}^{(1)}_{3/2}(\alpha_{jk}, \mathfrak{h}_j, \mathfrak{h}_k) \ , \tag{22c}$$

$$B_{jk} = \frac{n_j}{4} \left(\frac{m_k}{M_\odot + m_j} \right) \alpha_{jk} \tilde{b}^{(1)}_{3/2}(\alpha_{jk}, \mathfrak{h}_j, \mathfrak{h}_k), \quad j \neq k \ , \tag{22d}$$

that sum becomes

$$R_j = n_j a_j^2 \left[\tfrac{1}{2} A_{jj} (h_j^2 + k_j^2) + \sum_{k \neq j} A_{jk} (h_j h_k + k_j k_k) + \tfrac{1}{2} B_{jj} (p_j^2 + q_j^2) + \sum_{k \neq j} B_{jk} (p_j p_k + q_j q_k) \right] \tag{23}$$

in the notation of Murray & Dermott (1999). The A_{jk} and the B_{jk} can be regarded as two $N \times N$ matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ whose entries describe the magnitude of the mutual gravitational interactions that are exerted among the N rings. In the following discussion, quantities having a j subscript always refer to the perturbed ring in question, while the k subscript always refer to another perturbing ring.

The time variation of the rings' orbit elements is given by the Lagrange planetary equations; to lowest order in e and i, these are (Brouwer & Clemence 1961; Murray & Dermott 1999)

$$\frac{dh_j}{dt} \simeq \frac{\partial R_j / \partial k_j}{n_j a_j^2} = \sum_{k=1}^{N} A_{jk} k_k \ , \tag{24a}$$

$$\frac{dk_j}{dt} \simeq -\frac{\partial R_j / \partial h_j}{n_j a_j^2} = -\sum_{k=1}^{N} A_{jk} h_k \ , \tag{24b}$$

$$\frac{dp_j}{dt} \simeq \frac{\partial R_j / \partial q_j}{n_j a_j^2} = \sum_{k=1}^{N} B_{jk} q_k \ , \tag{24c}$$

$$\frac{dq_j}{dt} \simeq -\frac{\partial R_j / \partial p_j}{n_j a_j^2} = -\sum_{k=1}^{N} B_{jk} p_k \ , \tag{24d}$$

and their well-known Laplace-Lagrange solution is

$$h_j(t) = \sum_{i=1}^{N} E_{ji} \sin(g_i t + \beta_i) \ , \tag{25a}$$

$$k_j(t) = \sum_{i=1}^{N} E_{ji} \cos(g_i t + \beta_i) \ , \tag{25b}$$

$$p_j(t) = \sum_{i=1}^{N} I_{ji} \sin(f_i t + \gamma_i) \ , \tag{25c}$$

$$q_j(t) = \sum_{i=1}^{N} I_{ji} \cos(f_i t + \gamma_i) \ , \tag{25d}$$

where g_i is the ith eigenvalue of the $\boldsymbol{A}$ matrix, f_i is the ith eigenvalue of $\boldsymbol{B}$, E_{ji} is the $N \times N$ matrix formed from the N eigenvectors to $\boldsymbol{A}$, I_{ji} is the matrix of eigenvectors to $\boldsymbol{B}$, and β_i and γ_i are integration constants.

To apply this rings model, first assign to the N rings their masses m_j, their semimajor axes a_j, and their fractional half-thickness $\mathfrak{h}_j$. Planets are represented as thin rings having $\mathfrak{h}_j = 0$. Then construct the system's $\boldsymbol{A}$ and $\boldsymbol{B}$ matrices and compute their eigenvalues g_i and f_i, and the eigenvector arrays E_{ji} and I_{ji}. The rings' initial orbits $e_j(0)$, $i_j(0)$, $\tilde{\omega}_j(0)$, and $\Omega_j(0)$ are then used to determine the integration constants β_i and γ_i, as well as to rescale the eigenvectors E_{ji} and I_{ji} such that equations (25) agree with the initial conditions. A handy recipe for this particular task is given in Murray & Dermott (1999). Equations (25) are then used to compute the system's time history, and orbit elements are recovered via $e_j^2 = h_j^2 + k_j^2$, $i_j^2 = p_j^2 + q_j^2$, $\tan \tilde{\omega}_j = h_j/k_j$, and $\tan \Omega_j = p_j/q_j$.

The rather laborious derivation given above thus confirms the assertion by Tremaine (2001) that one only needs to soften the Laplace coefficients in order to use the Laplace-Lagrange solution for calculating the secular evolution of a continuous disk. However, § 2.2.1 shows that this softening must be done judiciously, such that equation (18a) is obeyed in order for the solution to preserve the system's angular momentum.

2.2. *Tests of the Rings Model*

Several tests have been devised in order to demonstrate that the rings model described above behaves as expected.

2.2.1. *Angular Momentum Conservation*

The equations developed above conserve the system's total z-component of angular momentum, $L_z = \sum_j m_j n_j a_j^2 (1 - e_j^2)^{1/2} \cos(i_j)$. To show this, expand L_z to second order in e' and i', which is the same degree of precision to which the disk potential is developed above. This gives $L_z \simeq L_0 - L_e - L_i$, where

$$L_e = \tfrac{1}{2} \sum_j m_j n_j a_j^2 e_j^2 \ , \tag{26a}$$

$$L_i = \tfrac{1}{2} \sum_j m_j n_j a_j^2 i_j^2 \ , \tag{26b}$$

and $L_0 = \sum_j m_j n_j a_j^2$ is a constant. We show in Appendix C that the dynamical equations (24) conserve angular momenta to second order in e and i, that is, $dL_e/dt = 0$ and $dL_i/dt = 0$. When the rings model is used to calculate the secular evolution of a "sparse" system, such as the Jupiter-Saturn system described below or one composed of all four giant planets, the angular momenta L_e and L_i are preserved to near machine limits to a fractional precision of $\sim 10^{-7}$ in this single floating-point implementation. This is true regardless of whether the planets are represented as thin rings having $\mathfrak{h}_j = 0$ or as thick rings with $\mathfrak{h}_j > 0$. However, L_z conservation is a little worse, $\sim 10^{-6}$, in the "crowded" systems described in § 4 that consist of a few planets plus a disk comprising many closely packed rings. These errors are always smaller, by a factor of $\sim 10^4$ to 10^5, than the angular momenta associated with the disturbances and waves seen in these disks. Thus, the wavelike disk behavior reported below is real and is not due to a diffusion of some numerical error.

2.2.2. *Jupiter and Saturn*

Murray & Dermott (1999) use the Laplace-Lagrange planetary solution, equations (25), to study a system composed of Jupiter and Saturn. The rings model developed here reproduces that system's $\boldsymbol{A}$ and $\boldsymbol{B}$ matrices, eigenvector arrays E_{ji} and I_{ji}, eigenvalues g_i and f_i, and integration constants β_i and γ_i, to the same precision quoted by Murray & Dermott (1999), provided one sets $n_j = (GM_\odot/a_j^3)^{1/2}$ and $\mathfrak{h}_j = 0$. The rings model also reproduces the figures in Murray & Dermott (1999) that show this system's orbital history, as well as the figures showing the forced orbit elements of numerous other massless "test rings" orbiting throughout this system. However, is should be noted that rigorous angular momentum conservation actually requires setting $n_j = [G(M_\odot + m_j)/a_j^3]^{1/2}$ instead.

2.2.3. *Precession in an Axisymmetric Disk*

Heppenheimer (1980) points out that a massless test particle orbiting in a smooth, axisymmetric disk experiences a *regression* of its longitude of periapse, i.e., $\dot{\tilde{\omega}} < 0$, which is the opposite of the usual prograde apse precession that occurs throughout the solar system. As Ward (1981) shows, it is the nearby disk parcels whose orbits actually cross the test particle's orbit that drive periapse regression at a rate that exceeds the prograde contribution from the more distant parts of the disk.

The rings code also reproduces this phenomenon. An annulus having a surface density σ', mass $\delta m' = 2\pi\sigma' a' da'$, a semimajor axis a', and a fractional half-thickness $\mathfrak{h}'$ contributes $\delta R = n^2 a^2 [f(\alpha, 0, \mathfrak{h}')e^2 - \alpha \tilde{b}_{3/2}^{(1)}(\alpha, 0, \mathfrak{h}')i^2]\delta m'/8\ M_\odot$ to the particle's disturbing function (see eq. [19] with $e' = i' = 0$). Adopting a power law in the disk surface density, $\sigma' = \sigma(a)\alpha^{-r}$ where $\alpha = a'/a$, the total disturbing function integrated across a semi-infinite disk is

$$R = \int \delta R = \tfrac{1}{2}\mu_d n^2 a^2 \left(I_{\tilde{\omega}} e^2 - I_\Omega i^2\right) , \tag{27}$$

where $\mu_d = \pi\sigma(a)a^2/M_\odot = \pi G\sigma/an^2$ is the so-called normalized disk mass and

$$I_{\tilde{\omega}} = \tfrac{1}{2}\int_0^\infty \alpha^{1-r} f(\alpha, 0, \mathfrak{h}')\, d\alpha , \tag{28a}$$

$$I_\Omega = \tfrac{1}{2}\int_0^\infty \alpha^{2-r} \tilde{b}_{3/2}^{(1)}(\alpha, 0, \mathfrak{h}')\, d\alpha . \tag{28b}$$

Note that $I_{\tilde{\omega}}$ and I_Ω are double integrals according to the definitions of $\tilde{b}_s^{(m)}$ and f, equations (12) and (17a). However, these integrals are analytic for selected power laws r. For instance, setting $r = 1$ or 2 and then instructing symbolic math software such as MAPLE to do the radial integration first and the angular integration second yields $I_{\tilde{\omega}} = -1/(1 + \mathfrak{h}'^2/2)^{1/2} \simeq -1$ and $I_\Omega = [\mathfrak{h}'^{-1} + (1 + \mathfrak{h}'^2/4)^{1/2} + \mathfrak{h}'/2]/[(1 + \mathfrak{h}'^2/4)(1 + \mathfrak{h}'^2/2)]^{1/2} \simeq 1/\mathfrak{h}'$ for small $\mathfrak{h}'$. The Lagrange planetary equations then give the test ring's precession rates:

$$\dot{\tilde{\omega}} \simeq \frac{\partial R/\partial e}{na^2 e} = I_{\tilde{\omega}}\mu_d n \simeq -\mu_d n , \tag{29a}$$

$$\dot{\Omega} \simeq \frac{\partial R/\partial i}{na^2 i} = -I_\Omega \mu_d n \simeq -\frac{\mu_d n}{\mathfrak{h}'} . \tag{29b}$$

Very similar precession rates were previously derived in Heppenheimer (1980) and Ward (1981). Note that the model rings start to overlap when their fractional radial thickness $2\mathfrak{h}'$ exceeds the rings' fractional separation $\delta = \Delta a/a$, so a massless test ring should precess at the above rates when the disk rings are sufficiently overlapping.

These expectations are tested by constructing a 50 $M_\oplus$ disk having an $r = 2$ power-law surface density using 200 circular, coplanar rings arranged over 10 AU $< a <$ 100 AU, with a number of thin, massless "test rings" also orbiting within this disk. Several simulations are performed with the massive rings having a variety of thicknesses $\mathfrak{h}'$. As expected, those test rings that reside far from the disk's edges precess at rates given by equation (29) whenever the disk rings are sufficiently overlapping, namely, when $\mathfrak{h}' \geq 2\delta$.

2.2.4. *Precession in an Eccentric Disk*

Although a particle's periapse will experience retrograde precession when embedded in an axisymmetric disk, prograde precession is possible in a nonaxisymmetric disk. For instance, prograde precession is evident in the N-body simulations of an eccentric stellar disk orbiting the putative black hole at the center of the Andromeda galaxy M31 (Jacobs & Sellwood 2001). These simulated disks have masses $\sim$0.1 times the central mass, with the interior parts of the disk being progressively more eccentric. These simulations reveal a long-lived overdense region in the direction of apoapse that persists because of a coherent alignment of the particles' periapse, with the density pattern rotating in a prograde sense at rates that increase with the disk mass. It should be noted that the disk particles' eccentricities are not always small, so these N-body simulations cannot be used as a quantitative benchmark for the rings code. Nonetheless, it is comforting to find that the rings code does indeed reproduce the density patterns seen in the N-body disks that precess at rates very similar to that reported in Jacobs & Sellwood (2001).

3. SPIRAL WAVE THEORY

Section 4 uses the preceding rings model to demonstrate that apsidal density waves and nodal bending waves may have once propagated throughout the Kuiper Belt. An apsidal wave is a one-armed spiral density wave that slowly rotates over a periapse precession timescale. Similarly, a nodal wave is a one-armed spiral bending wave that rotates over a nodal precession timescale.

A brief review of spiral wave theory is in order. Many of the waves' properties, such as their wavelength and propagation speed, are readily extracted from the waves' dispersion relations. These dispersion relations are usually obtained from solutions to the Poisson and Euler equations for the disk. However, the following discussion will show that these dispersion relations can also be derived from the Lagrange planetary equations.

3.1. *Apsidal Density Waves*

The disturbing function $R(a)$ for the disk material orbiting at a semimajor axis a is obtained from equation (19) with the perturbing mass m' replaced by the differential mass dm', whose contributions are integrated across a semi-infinite

disk:

$$R(a) = \tfrac{1}{2}\mu_d n^2 a^2 \times \int_0^\infty \alpha^{1-r}\left[\tfrac{1}{2}e^2 f(\alpha, \mathfrak{h}, \mathfrak{h}) + ee' \cos(\tilde{\omega}' - \tilde{\omega}) g(\alpha, \mathfrak{h}, \mathfrak{h})\right] d\alpha \ . \quad (30)$$

where $e'(\alpha)$ and $\tilde{\omega}'(\alpha)$ are the orbit elements of the perturbing parts of the disk, the unprimed quantities refer to the perturbed annulus at a, r is the power law for the disk's surface density variation, and a constant fractional thickness $\mathfrak{h}$ is assumed throughout the disk. The Lagrange planetary equations then give the disk's periapse precession rate at a:

$$\dot{\tilde{\omega}}(a) \simeq \frac{\partial R/\partial e}{na^2 e} = \mu_d n (I_{\tilde{\omega}} + I_{\rm dw}) \ , \quad (31)$$

where $I_{\tilde{\omega}} \simeq -1$ is from the left-hand term in equation (30) and is the contribution by an undisturbed disk to its own precession (see eqs. [28]–[29]), and the right-hand term is

$$I_{\rm dw} = \frac{1}{2}\int_0^\infty \frac{e'}{e}\cos(\tilde{\omega}' - \tilde{\omega})\alpha^{1-r} g(\alpha, \mathfrak{h}, \mathfrak{h}) d\alpha \ , \quad (32)$$

which is the relative rate at which the density wave drives its own precession.

It is expected that the eccentricities associated with a spiral density wave will vary only slowly with distance a, such that $e'(\alpha)/e \simeq 1$. The spiral wave will also organize the disk's longitude of periapse $\tilde{\omega}$ such that it varies as $\tilde{\omega}(a, t) = \tilde{\omega}t - \int^a k(A)dA$, where $k(a)$ is the wavenumber and $\lambda = 2\pi/|k|$ is the radial wavelength. If the spiral wave pattern is tightly wound such that $\lambda \ll a$ and $|ka| \gg 1$, then the dominant contributions to $I_{\rm dw}$ are largely due to the nearby parts of the disk where $\alpha = a'/a \sim 1 \pm \lambda/a$ and $\tilde{\omega}' - \tilde{\omega} = -\int_a^{a'} k(A)dA \simeq -k(a' - a)$, while the contributions from the more distant parts of the disk tend to cancel owing to the rapid oscillation of the cosine factor. Thus we can set $\alpha = 1$ in equation (32) except where it appears as the combination $x = \alpha - 1$ where $|x| \ll 1$. In this case the softened Laplace coefficients that are present in the g function can be replaced with the approximate forms that are valid for $|x| \ll 1$ (see eq. [B5]), so g becomes

$$g(x) \simeq \frac{2}{\pi}\frac{2\mathfrak{h}^2 - x^2}{(2\mathfrak{h}^2 + x^2)^2} \ . \quad (33)$$

It is also permitted to extend the lower integration limit in equation (32) to $-\infty$ in the tight-winding limit, so

$$I_{\rm dw} \simeq \tfrac{1}{2}\int_{-\infty}^{\infty} \cos(|ka|x) g(x)\, dx = |ka| e^{-\sqrt{2}\mathfrak{h}|ka|} \ . \quad (34)$$

And finally, if the wave is to remain coherent across this disk, this self-precession must occur at the same constant rate ω throughout the disk, so $\dot{\tilde{\omega}}(a) = \omega \simeq \mu_d(|ka|e^{-\sqrt{2}\mathfrak{h}|ka|} - 1)n$. Note that this disturbing frequency ω, which is also called the pattern speed, can also be identified as any one of the eigenfrequencies g_i that appear in equations (25). Usually it is another perturber that is responsible for launching the wave and causing the disk to precess in concert at the rate ω, and this is usually at a rate that dominates over the nonwave contribution to the disk's precession, i.e., $|\omega| \gg \mu_d n$. This then yields the dispersion relation for tightly wound apsidal density waves:

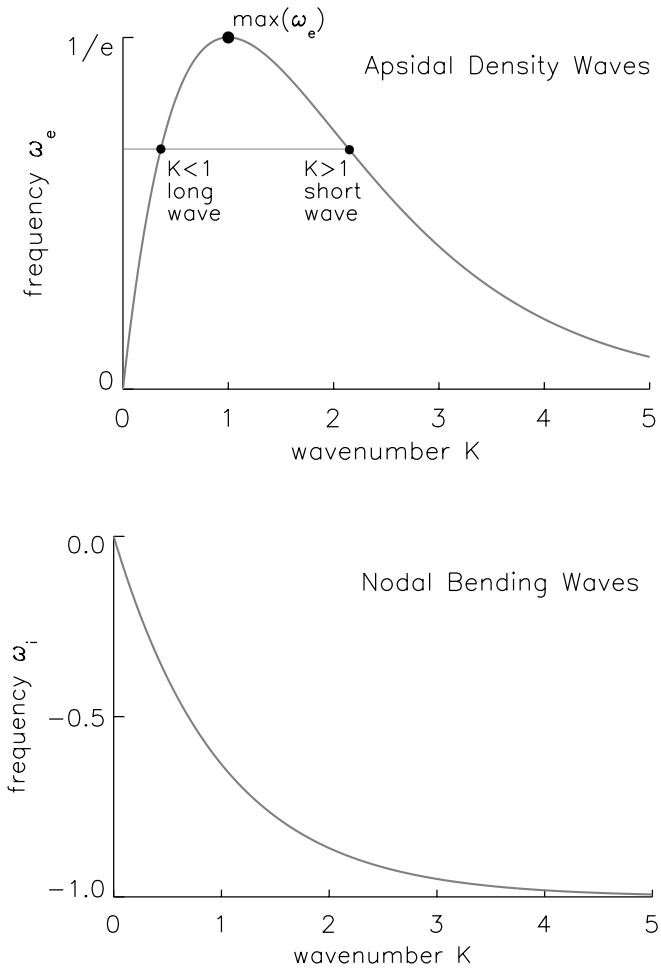

FIG. 2.—*Top*: Dispersion relation $\omega_e = Ke^{-K}$ for apsidal density waves. *Bottom*: Dispersion relation $\omega_i = e^{-K} - 1$ for nodal bending waves.

$$\omega \simeq e^{-\sqrt{2}\mathfrak{h}|ka|}\mu_d |ka| n \ , \quad (35)$$

which has the dimensionless form

$$\omega_e(K) = Ke^{-K} \ , \quad (36)$$

where $\omega_e = \sqrt{2}\mathfrak{h}\omega/\mu_d n$ is the dimensionless frequency and $K = \sqrt{2}\mathfrak{h}|ka|$ is the dimensionless wavelength; Figure 2 plots ω_e versus K. A more general dispersion relation for an m-armed spiral wave in a stellar disk is given in Toomre (1969), and in the limit that $|\omega| \ll n$ the resulting formula with $m = 1$ behaves qualitatively quite similar[4] to equation (36). Note that equation (35) also recovers the usual dispersion relation $\omega = \mu_d |ka| n$ for apsidal waves in an infinitesimally thin disk when $\mathfrak{h} = 0$.

Note that $\omega_e(K) > 0$ and that it also has a maximum at $K = 1$ where $\omega_e(1) = \exp(-1)$. Since ω_e is a function of semimajor axis a, the restriction $0 < \omega_e(a) \lesssim 0.368$ indicates that apsidal waves can only propagate in a restricted

[4] In this limit, Toomre's dispersion relation becomes $\omega_T(K) = K\mathscr{F}$ where the reduction factor $\mathscr{F}$ is a more complicated function of K. However, a numerical evaluation of this function shows that $\omega_T(K)$ has the same form as the $\omega_e(K)$ curve shown in Fig. 2.

interval in a. This restriction can also be viewed as a constraint on the disk thickness, namely,

$$\mathfrak{h} \lesssim 0.260\mu_d|n/\omega| \equiv \mathfrak{h}_Q \ . \tag{37}$$

Alternatively, equation (37) can also be viewed as an upper limit on the frequency ω or a lower limit on the disk mass μ_d wherein wave action is permitted.

Waves having a wavenumber $K < 1$ are called long waves, since they have a wavenumber $|k_L| \simeq (\omega/n)/\mu_d a$ and the longer wavelength $\lambda_L = 2\pi/|k_L| \simeq 2\pi\mu_d(n/\omega)a$, while short waves have wavenumbers $K > 1$ or $|k_S| > 1/\sqrt{2}\mathfrak{h}a$ and the shorter wavelength $\lambda_S = 2\pi/|k_S| < 2\sqrt{2}\pi\mathfrak{h}a$. These long and short density waves also correspond to the g and p modes, respectively, of Tremaine (2001). Writing λ_L in terms of $\mathfrak{h}_Q$ and then requiring $\mathfrak{h} < \mathfrak{h}_Q$ also means that apsidal waves can propagate wherever $\lambda_L \gtrsim 24h$.

The rate at which apsidal waves propagate across the disk is given by their group velocity

$$c_g = \frac{d\omega}{dk} = s_k\mu_d(1-K)e^{-K}na \ , \tag{38}$$

where $s_k = \mathrm{sgn}\,(k)$. Waves with $s_k = +1$ are called trailing waves, and their longitude of perihelia $\tilde{\omega}$ decreases with increasing semimajor axis a, while $s_k = -1$ are leading waves whose longitude increases with a. Since the waves' group velocity c_g is proportional to the slope of the $\omega_e(K)$ curve, the site where $K = 1$ is a turning point where wave reflection occurs; in galactic dynamics this reflection site is known as a Q barrier. A long wave that approaches the Q barrier from the $K < 1$ side of Figure 2 thus reflects as a short wave as it continues along the $K > 1$ side of the curve. The simulations shown in § 4 also show that long trailing waves that instead strike a disk edge also reflect as short trailing waves.

3.2. *Surface Density Variations*

The compression or rarefaction among the disk's rings, or streamlines, is $(\partial r/\partial a)^{-1}$, which is also the relative change in the disk's surface density σ/σ_0 associated with a density wave (e.g., Borderies, Goldreich, & Tremaine 1985). For a small amount of compression, $\sigma/\sigma_0 = 1 + \Delta\sigma/\sigma_0$, where the fractional surface density variation is

$$\frac{\Delta\sigma}{\sigma_0} = \left(\frac{\partial r}{\partial a}\right)^{-1} - 1 \simeq \frac{\partial(ea)}{\partial a}\cos(\phi - \tilde{\omega}) + ea\frac{\partial\tilde{\omega}}{\partial a}\sin(\phi - \tilde{\omega}) \tag{39}$$

to lowest order in e. The second term dominates over the first in the tight-winding limit, so $|\Delta\sigma/\sigma_0| \sim O(e|ka|)$. Density waves are nonlinear when $|\Delta\sigma/\sigma_0| > 1$, and these large density variations are a consequence of overlapping streamlines. For long density waves having a wavenumber $|k_L| = \omega/\mu_d an$, the disk's streamlines will cross when the waves' eccentricities exceed $e_L \sim \mu_d\omega/n$, while streamline crossing occurs among short waves when eccentricities exceed $e_S \sim \sqrt{2}\mathfrak{h}$. As the simulations of § 4 show, a dynamically cool Kuiper Belt is very susceptible to the propagation of short nonlinear density waves that facilitate streamline crossing. Depending on the relative velocities of these crossed streamlines, apsidal wave action might either encourage accretion or enhance collisional erosion among KBOs.

3.3. *Nodal Bending Waves*

The derivation of the dispersion relation for nodal bending waves, and its analysis, proceeds similarly. The disk's integrated disturbing function is

$$R(a) = -\tfrac{1}{2}\mu_d n^2 a^2 \times \int_0^\infty \alpha^{2-r}\tilde{b}^{(1)}_{3/2}(\alpha,\mathfrak{h},\mathfrak{h})\left[\tfrac{1}{2}i^2 - ii'\cos(\Omega' - \Omega)\right]d\alpha \ , \tag{40}$$

so the Lagrange planetary equation gives

$$\dot{\Omega}(a) \simeq \frac{\partial R/\partial i}{na^2 i} = \mu_d n(I_{\rm bw} - I_\Omega) \ , \tag{41}$$

where $\Omega(a,t) = \dot{\Omega}t - \int^a k(A)\,dA$, and

$$I_{\rm bw} = \frac{1}{2}\int_0^\infty \alpha^{2-r}\tilde{b}^{(1)}_{3/2}(\alpha,\mathfrak{h},\mathfrak{h})\cos[ka(\alpha-1)]\,d\alpha \simeq \frac{e^{-\sqrt{2}\mathfrak{h}|ka|}}{\sqrt{2}\mathfrak{h}} \tag{42}$$

is the bending wave's contribution to its own precession. The contribution from the undisturbed disk is $I_\Omega \simeq 1/\sqrt{2}\mathfrak{h}$, where the additional $\sqrt{2}$ factor is the result of changing the middle argument in equation (28b) from 0 to $\mathfrak{h}$. Since $\omega = \dot{\Omega}$ is a constant for a coherent wave, the dispersion relation for tightly wound nodal bending waves is

$$\omega \simeq \frac{\mu_d}{\sqrt{2}\mathfrak{h}}\left(e^{-\sqrt{2}\mathfrak{h}|ka|} - 1\right)n \ . \tag{43}$$

Note that the usual dispersion relation for nodal waves in an infinitesimally thin disk, $\omega \simeq -\mu_d|ka|n$, is obtained when $\mathfrak{h} \to 0$.

The dimensionless form of the dispersion relation is

$$\omega_i(K) = e^{-K} - 1 \ , \tag{44}$$

where $\omega_i = \sqrt{2}\mathfrak{h}\omega/\mu_d n$, and is plotted in Figure 2. As the figure shows, nodal bending waves can propagate only in regions where $-1 < \omega_i(a) < 0$, which similarly limits the disk thickness to $\mathfrak{h} \lesssim 2.72\mathfrak{h}_Q$. The nodal waves' group velocity is

$$c_g = \frac{d\omega}{dk} = -s_k\mu_d(1+\omega_i)na \ , \tag{45}$$

which indicates that nodal waves tend to stall, i.e., $|c_g| \to 0$ as they approach the $\omega_i = -1$ boundary. Note that this dispersion relation only admits a long-wavelength solution having a wavenumber $|k_L| \simeq -\omega/\mu_d na$ and a wavelength $\lambda_L \simeq 2\pi\mu_d|n/\omega|a$ for waves far from the stall zone. Since $\mathfrak{h} \lesssim 2.72\mathfrak{h}_Q$, the disk can sustain nodal waves wherever $\lambda_L \gtrsim 9h$. But if an outward-traveling long leading wave with $s_k = -1$ encounters a disk edge, it will reflect as an $s_k = +1$ long trailing wave. Examples of nodal wave reflection and stalling are also given below.

4. THE SECULAR EVOLUTION OF THE PRIMORDIAL KUIPER BELT

Using the recipe given in § 2.1.1, the rings model is used to compute the secular evolution of the primordial Kuiper Belt as it is perturbed by the four giant planets. In these

simulations the giant planets are represented by thin $\mathfrak{h} = 0$ rings whose initial orbits are their current orbits, while the Kuiper Belt is represented by 500 rings whose semimajor axes extend from 36 AU out to 50–70 AU. The location of the belt's inner edge is chosen such that only the outermost radial and vertical secular resonances, the ν_8 and the ν_{18}, reside in this disk near 40 AU when of low mass. The semimajor axes of each belt ring increases as $a_{j+1} = (1 + \delta)a_j$ where the rings' fractional separation δ is typically $\sim$0.001. The rings' fractional half-thickness $\mathfrak{h}$ is always in excess of 2δ, as is required to get the correct apse precession in an axisymmetric disk (see § 2.2.3). The belt rings' initial eccentricities and inclinations are zero, with all inclinations being measured with respect to the system's invariable plane. The mass of each ring is chosen such that the belt's surface density $\sigma(a)$ varies as $a^{-1.5}$. For this configuration, if the total Kuiper Belt mass over the 36 AU $\leq a \leq$ 70 AU zone is $M_{\rm total}$, then the total mass in the "observable" 30 AU $\leq a \leq$ 50 AU zone that is currently accessible to astronomers would be $M_{\rm KB} = 0.67 M_{\rm total}$ had the above surface density law extended inward to 30 AU. For these systems the normalized disk mass is $\mu_d \simeq M_{\rm KB}/M_\odot$.

4.1. *A $M_{KB} = 10\ M_\oplus$ Example*

Figure 3 shows a snapshot of apsidal density waves as they propagate across an $M_{\rm KB} = 10\ M_\oplus$ Kuiper Belt having a half-thickness $\mathfrak{h} = 5\delta = 0.0067$. Since $\mathfrak{h} \sim 0.2\mathfrak{h}_Q$, the necessary disk conditions for the propagation of apsidal and nodal waves are well satisfied. Initially, a long trailing density wave is launched at the belt's inner edge. This wave is really more like a pulse $\sim$5 AU wide, and Figure 3 shows that by time $t = 2 \times 10^6$ yr the wave has just started to reflect at the disk's outer edge at 70 AU. The gray-scale map of the disk's surface density variation, $\Delta\sigma/\sigma$, is obtained using equation (39), and this map also provides a historical record of this system's wave action. It should be noted that equation (39) is quantitatively correct only when $|\Delta\sigma/\sigma| \ll 1$, a condition that is rarely satisfied by the results obtained here. Nonetheless, equation (39) is still useful in a qualitative sense since it will indicate the disk regions where large surface density variations as well as orbit crossing can be expected. As the outer edge of the $\Delta\sigma/\sigma$ map shows, the main apsidal density wave pulse at $67 < a < 70$ AU has just reflected at a short trailing wave, and this nonlinear wave, having $|\Delta\sigma/\sigma| > 1$, will completely dominate the disk's appearance at later times as it propagates inward. But until this happens, the bulk of the disk's appearance over $45 < a < 67$ AU is still dominated by lower amplitude long waves that are following behind the main density pulse. Note also that short leading waves were also emitted at the disk's inner edge, but as the density gray scale shows, they have only propagated out to $a = 45$ AU thus far owing to their slower group velocity (see eq. [38]). These short waves are well resolved in the sense that their radial wavelength of $\lambda_S \sim 1$ AU spans about 15 disk rings. Although these short waves are seen in the $e(a)$ and $\tilde{\omega}(a)$ plots as only tiny wiggles over 36 AU $< a <$ 45 AU that are superimposed on top of the disk's longer wavelength behavior, the density gray scale shows that the short waves dominate the inner disk's appearance. The electronic edition of Figure 3 is also linked to an animated sequence of these figures that give the system's complete time history. That animation shows that by time $t \simeq 1 \times 10^7$ yr, nonlinear short waves will have swept across the entire disk, and they result in large surface density variations $|\Delta\sigma/\sigma| > 1$ over radial wavelengths of $\lambda_S \sim 1$ AU.

Figure 3 also shows that by time $t = 2 \times 10^6$ yr, a long leading nodal bending wave pulse has already propagated across the disk, where it has reflected at the disk's outer edge and just started to return inward as a long trailing bending wave. But when that pulse reaches the disk's inner edge, a portion of the wave's angular momentum content will continue to propagate farther inward, where it will give a small kick to the giant planets' orbit elements, while the remaining wave pulse reflects again and propagates outward. The same phenomenon also occurs among the apsidal density waves. Thus, after a few reflections, a single wave pulse will lose its initial spatial coherence by spawning multiple wave trains that, in this friction-free model, forever roam about the belt. This ultimately results in a rather wobbly-looking standing density wave pattern that varies over the short-wavelength scale $\lambda_S \sim 1$ AU, as well as a standing bending wave pattern that varies over a much longer scale $\lambda_L \sim 40$ AU.

A dynamical "spectrum" of this Kuiper Belt is shown in Figure 4, which plots all of the eccentricity eigenvector elements $|E_{ji}|$ for each of the disk rings versus their eigenfrequency g_i, as well as inclination eigenvector elements $|I_{ji}|$ versus eigenfrequency $|f_i|$. The upper figure is quite reminiscent of the findings of Tremaine (2001), who showed that a gravitating disk tends to exhibit its strongest response to slow radial perturbations via modes having discrete patterns speeds ω that can be identified with any of the peak frequencies g_i seen in Figure 4. Figures such as these are quite useful for identifying the patterns speeds ω that are associated with the density and bending waves that propagate in the disk.

4.2. *Variations with Kuiper Belt Mass M_{KB}*

Simulations have been performed with Kuiper Belts having masses $M_{\rm KB} = 0$, 0.08, 0.2, 2, 10, and 30 $M_\oplus$ in the observable 30 AU $\leq a \leq$ 50 AU zone [again, this would be the belt's mass assuming its surface density $\sigma(a) \propto a^{-1.5}$ were to extend solely over that region]. The results are summarized in Figure 1, which shows the maximum eccentricities $e_{\rm max}$ and maximum inclinations $i_{\rm max}$ achieved by the rings in each simulation. The belt's radial width is indicated by the breadth of the curves in Figure 1, which ranges from 34 AU in the higher mass belts to 14 AU for the $M_{\rm KB} = 0.08\ M_\oplus$ system. Each simulation uses 500 disk rings having a fractional half-thickness $\mathfrak{h} = 2\delta$, so the three higher mass systems have $\mathfrak{h} = 0.0027$, while the lower mass disk $M_{\rm KB} = 0.2\ M_\oplus$ is somewhat thinner with $\mathfrak{h} = 0.0015$, and the $M_{\rm KB} = 0.08\ M_\oplus$ system has $\mathfrak{h} = 0.0010$. The rings in these lower mass disks are more closely packed, so that their shorter wavelength density waves are well resolved, and they are also made thinner so as to push their Q barriers a bit further downstream. The disk's initial orbits are $e = 0 = i$, except for the $M_{\rm KB} = 0$ system, which instead adopts the forced orbit elements appropriate for a massless disk (e.g., Brouwer & Clemence 1961; Murray & Dermott 1999). These systems are evolved until their initial density and bending wave pulses have had the opportunity to reflect multiple times and have largely dissolved into standing wave patterns. The lower mass disks necessarily have longer run times because of their waves' slower propagation speeds (see eq. [38]); these run times are 1×10^9, 2×10^9, 1×10^8,

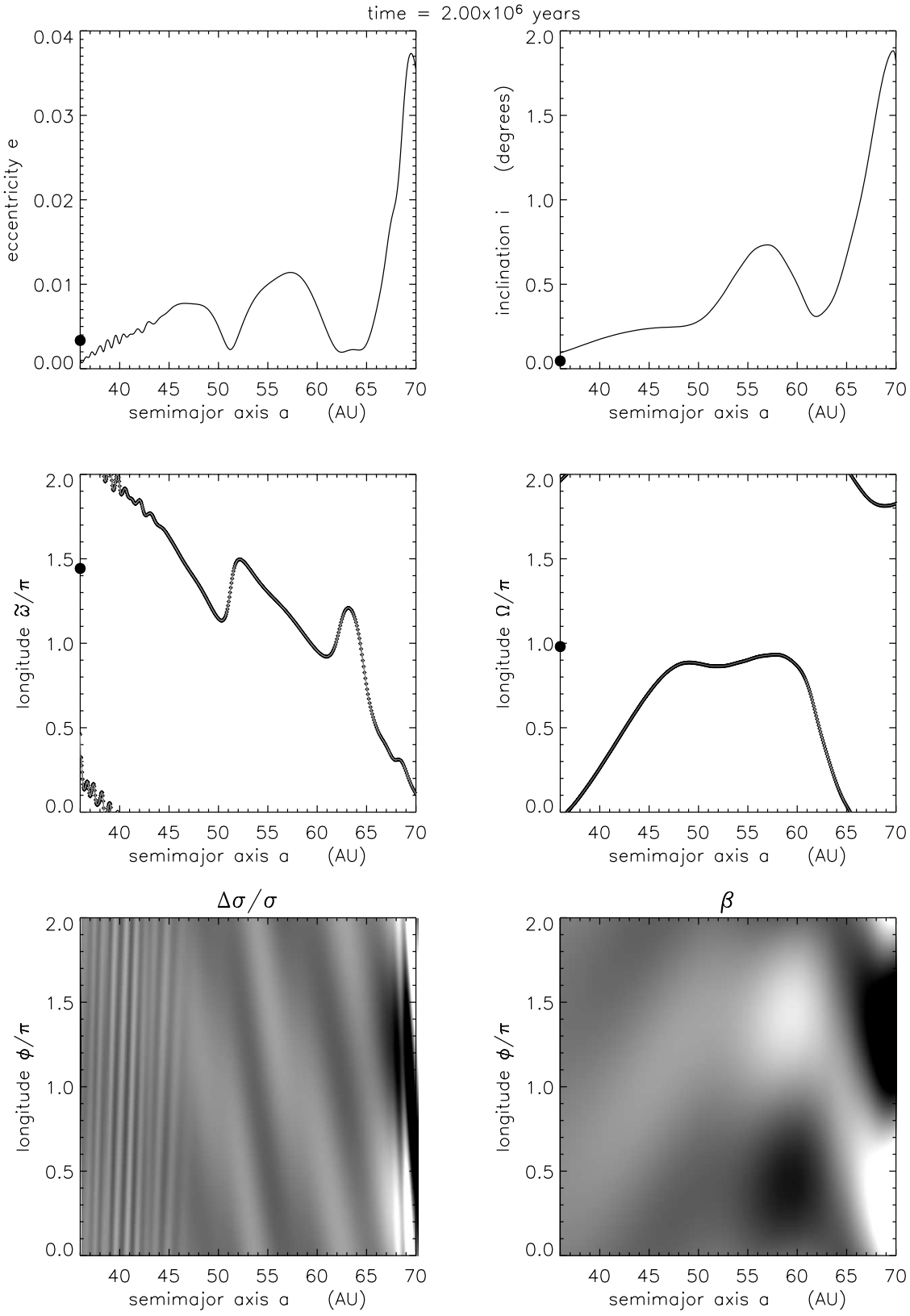

FIG. 3.—Snapshot of the simulated Kuiper Belt's orbit elements $(e, i, \tilde{\omega}, \Omega)$ plotted vs. semimajor axis a at time $t = 2 \times 10^6$ yr in a $M_{\rm KB} = 10\ M_\oplus$ belt. The belt's fractional half-thickness is $\mathfrak{h} = 0.0067$. The dots along the left axes indicate Neptune's orbit elements. The $\Delta\sigma/\sigma$ gray scale shows the disk's fractional surface density variations vs. the polar coordinates (a, ϕ), estimated via eq. (39). White/black zones indicate over/underdense regions where $|\Delta\sigma/\sigma|$ exceeds 0.63, while white/black zones in the other gray scale indicate the disk's latitude $\beta = z/a$ above/below the invariable plane, with saturation occurring wherever $|\beta|$ exceeds 0°.86. This figure is also available as an mpeg animation in the electronic edition of the Astrophysical Journal.

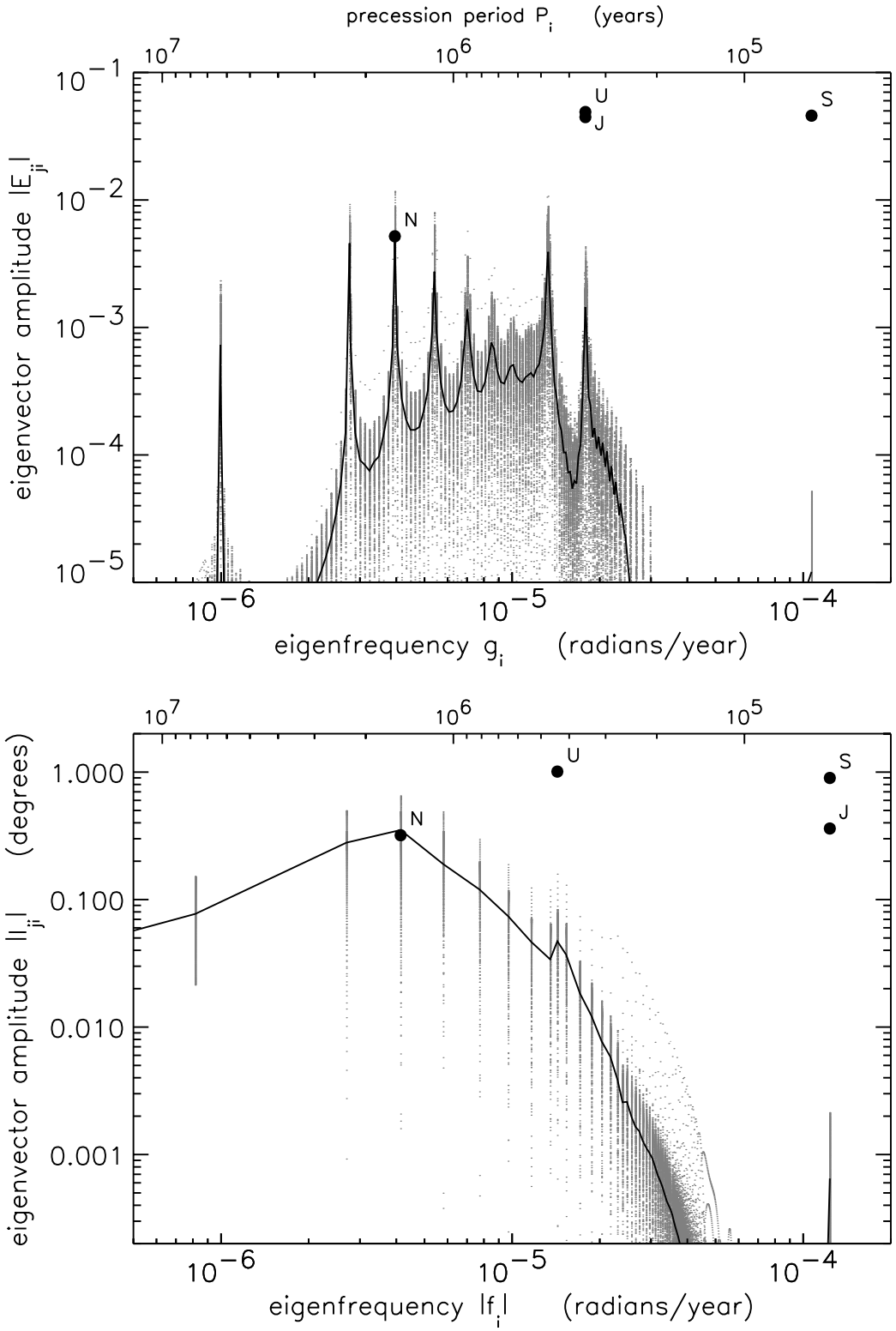

FIG. 4.—Numerous small dots show the individual elements in the disk rings' eccentricity eigenvectors $|E_{ji}|$, all plotted vs. their eigenfrequencies g_i, as well as the disk rings inclination eigenvector elements $|I_{ji}|$ vs. $|f_i|$, for the system shown in Fig. 3. The solid curves are $|E_{ji}|$ and $|I_{ji}|$ averaged over semimajor axes 45 AU $\leq a_j \leq$ 55 AU. The upper horizontal axes are the corresponding precession periods $2\pi/g_i$ and $2\pi/|f_i|$. The large filled circles indicate the eigenfrequency and eigenvector element that dominates the motion of each giant planet: J = Jupiter, S = Saturn, U = Uranus, and N = Neptune.

2×10^7, and 1×10^7 yr for the $M_{\rm KB} = 0.08$, 0.2, 2, 10, and 30 $M_\oplus$ systems, respectively. Once the standing wave pattern has emerged, the rings' instantaneous eccentricities and inclinations range over $0 \lesssim e < e_{\rm max}$ and $0 \lesssim i < i_{\rm max}$.

As Figure 1 shows, there are no peaks in the $e_{\rm max}$ and $i_{\rm max}$ curves for the higher mass disks having $M_{\rm KB} \geq 2\ M_\oplus$, indicating that there are no secular resonances in the disk itself; any such resonances likely lie between the orbits of Neptune

and the disk's inner edge at 36 AU. The simulations of these higher mass disks show long trailing density waves and long leading bending waves being launched at the disk's inner edge. These waves sweep across the disk, reflect at the disk's outer edge, and return as short trailing density waves and long trailing bending waves, similar to the history described in § 4.1 and seen in Figure 3.

However, secular resonances at $a \simeq 41$ AU are quite prominent in the lower mass disks having $M_{\rm KB} = 0.08$ and $0.2\ M_\oplus$. These are sites that launch long density and bending waves, and as Figure 1 shows, the density waves are able to propagate out as far as $a = 45$ and 49 AU, respectively, where they are reflected by a Q barrier and return as short density waves. These reflection sites occur where $\mathfrak{h} \simeq 2.2\mathfrak{h}_Q$ and $\mathfrak{h} \simeq 1.4\mathfrak{h}_Q$, which shows that equation (37) is an approximate yet useful indicator of where apsidal waves are allowed to propagate. Note also the large amplitudes of the density waves in these low-mass disks, $0.3 \lesssim e_{\rm max} \lesssim 1$, which clearly violates the model's assumption of low eccentricities. Thus, these particular curves should not be taken at face value. Nonetheless, they do indicate that apsidal waves in a low-mass Kuiper Belt may result in large eccentricities, and Figure 1 shows that sizable inclinations can also result from nodal bending waves. Indeed, in the $M_{\rm KB} = 0.08\ M_\oplus$ disk, maximum inclinations are typically $i \sim 20°$, which is comparable to the Main Belt's high-inclination component.

The larger e and i seen in the lower mass disks are a consequence of the giant planets transmitting a small fraction of their initial angular momentum deficit[5] (AMD) into the disk in each simulation. In each simulation, the giant planets deposit $\sim 0.005 L_e$ and $\sim 0.1 L_i$ into the disk's inner edge, which waves then transport and smear out across a vast swath of the Kuiper Belt. Since the AMD deposited in the disk is roughly constant in each simulation, the lower mass disks exhibit larger e and i excitations. It should also be noted that computational limitations in dynamical studies of the Kuiper Belt, particularly N-body simulations, often require treating the belt as a swarm of massless test particles. However, the comparison of the $M_{\rm KB} = 0$ curve to the $M_{\rm KB} > 0$ curves in Figure 1 shows that the end state of a Kuiper Belt having even just a modest amount of mass can be radically different from one naively treated as massless.

4.3. *Variations with Disk Thickness*

Figure 5 shows how the response of an $M_{\rm KB} = 10\ M_\oplus$ belt varies with increasing disk thickness $\mathfrak{h} = (2,\ 20,\ 30,\ 60,\ 100)\delta = (0.0027, 0.027, 0.04, 0.08, 0.13)$. According to equation (37), $\mathfrak{h}_Q(a) \propto a^{1/2-r=-1}$ in these $r = 3/2$ disks, so the Q barrier will move inward as the disk's $\mathfrak{h}$ is increased, as is evident in Figure 5. All of these disks have a normalized disk mass $\mu_d \simeq 3 \times 10^{-5}$, a mean motion $n \simeq 0.02$ rad yr^{-1}, and a pattern speed that is typically $\omega \sim 3 \times 10^{-6}$ rad yr^{-1}, so wave action is shut off when the disk thickness $\mathfrak{h}$ exceeds $\mathfrak{h}_Q = 0.26\mu_d|n/\omega| \sim 0.05$. In the thinnest disk with $\mathfrak{h} = 0.0027$, the Q barrier lies beyond the disk's outer edge at 70 AU, so long and then short apsidal density waves are able to slosh about the disk's full extent. However, the Q barrier lies in the disk at $a \simeq 60$ AU when $\mathfrak{h} = 0.027$ (Fig. 5, *red curve*), at $a \simeq 53$ AU when $\mathfrak{h} = 0.04$ (*green curve*), and apsidal waves are prohibited in the disks with $\mathfrak{h} \geq 0.08$ (*blue and purple curves*).

[5] For example, the L_e and L_i of eq. (26).

Figure 5 also shows that the nodal bending waves are shut off when the disk thickness $\mathfrak{h}$ exceeds the somewhat more relaxed criterion $2.72\mathfrak{h}_Q \sim 0.14$. Note also the peak at $a \simeq 53$ AU in the $\mathfrak{h} = 0.027$ disk (*red curve*) and at $a \simeq 39$ AU in the $\mathfrak{h} = 0.04$ disk (*green curve*). These particular disks admit two spiral patterns, a higher amplitude spiral that corotates with Neptune's dominant eigenmode at the pattern speed of $\omega \sim -3 \times 10^{-6}$ rad yr^{-1}, and a lower amplitude mode that corotates with Uranus at the faster rate $\omega \sim -1.5 \times 10^{-5}$ rad yr^{-1}. Since $\mathfrak{h}_Q \propto |\omega|^{-1}$, the faster spiral pattern has $2.72\mathfrak{h}_Q \sim 0.03$, which is why this wave stalls at $a = 53(39)$ AU in the $\mathfrak{h} = 0.027(0.04)$ disks, while the slower spiral mode is able to propagate the full breadth of these disks.

The behavior of a lower mass disk with $M_{\rm KB} = 0.2\ M_\oplus$ is shown in Figure 6 for disks having $\mathfrak{h} = (2,\ 5,\ 10,\ 15)\delta = (0.0015, 0.0037, 0.0074, 0.011)$. A pair of secular resonances lie near $a \simeq 40$ AU, and they launch apsidal and nodal waves having pattern speeds $\omega \sim \pm 3 \times 10^{-6}$ rad yr^{-1}. These disks have $\mu_d \simeq 6 \times 10^{-7}$ and $\mathfrak{h}_Q \sim 0.001$, and as the upper plot shows, further increases in $\mathfrak{h}$ bring the Q barrier inward until the wave-propagation zone has shrunk down to zero. The lower plot also shows that nodal waves forever slosh about the model Kuiper Belt in the $\mathfrak{h} = 0.0015$ disk (*orange curve*), whereas the peaks in the other curves show that these waves stall at sites ever closer to resonance in progressively thicker disks. These figures also show that when $\mathfrak{h} \gg \mathfrak{h}_Q$, the motions of a nongravitating disk are recovered, namely, the disk's maximum e and i are twice the forced motions seen in the $M_{\rm KB} = 0$ disk (*black curves*), with the factor of 2 being a consequence of these disks' initial conditions $e = 0 = i$.

4.4. *Implications for the Primordial Kuiper Belt*

The primordial Kuiper Belt likely had an initial mass of $M_{\rm KB} \sim 30\ M_\oplus$ (see § 1), and accretion models show that the initial KBO swarm must have had dispersion velocities less than ~ 0.001 Keplerian (Kenyon & Luu 1999), so $\mathfrak{h} \lesssim 0.001$, $\mu_d \sim 1 \times 10^{-4}$, and thus $\mathfrak{h}_Q \sim 0.2$, assuming the spiral waves have pattern speeds of $|\omega| \sim 3 \times 10^{-6}$ rad yr^{-1}. Since $\mathfrak{h} < \mathfrak{h}_Q$, the primordial Kuiper Belt readily sustained apsidal and nodal waves. Figure 1 shows that in this high-mass environment, these will be rather low amplitude waves having $e \sim \sin i \sim 0.01$. These waves will quickly propagate across a Kuiper Belt of width Δa in time $T_{\rm prop} \sim \Delta a/|c_g| \simeq \Delta a/\mu_d an$, so the wave propagation time is

$$T_{\rm prop} \sim 3 \times 10^5 \left(\frac{M_{\rm KB}}{30\ M_\oplus}\right)^{-1} \left(\frac{\Delta a}{30\ {\rm AU}}\right)\ {\rm yr}\ . \qquad (46)$$

In the friction-free model employed here, the outward-bound long density waves eventually reflect, either at a Q barrier in the disk (which might lie downstream where $\mathfrak{h} = \mathfrak{h}_Q$) or else at the disk's outer edge. The reflected waves then propagate inward as short density waves, and such waves are nonlinear in the sense that their surface density variations $\Delta\sigma/\sigma$ typically exceed unity. Figure 3 shows a snapshot of long and short density waves in a $M_{\rm KB} = 10\ M_\oplus$ disk. Note that the long waves completely dominate the disk's orbit elements $e(a)$ and $\tilde{\omega}(a)$ that vary over a wavelength of $\lambda_L \sim 10$ AU, while the short waves are the tiny variations in $e(a)$ and $\tilde{\omega}(a)$ that occur over a $\lambda_S \sim 1$ AU scale at the disk's inner and outer edge. Even though the short waves are almost imperceptible in the orbit-elements plots, the gray-scale map shows that these nonlinear

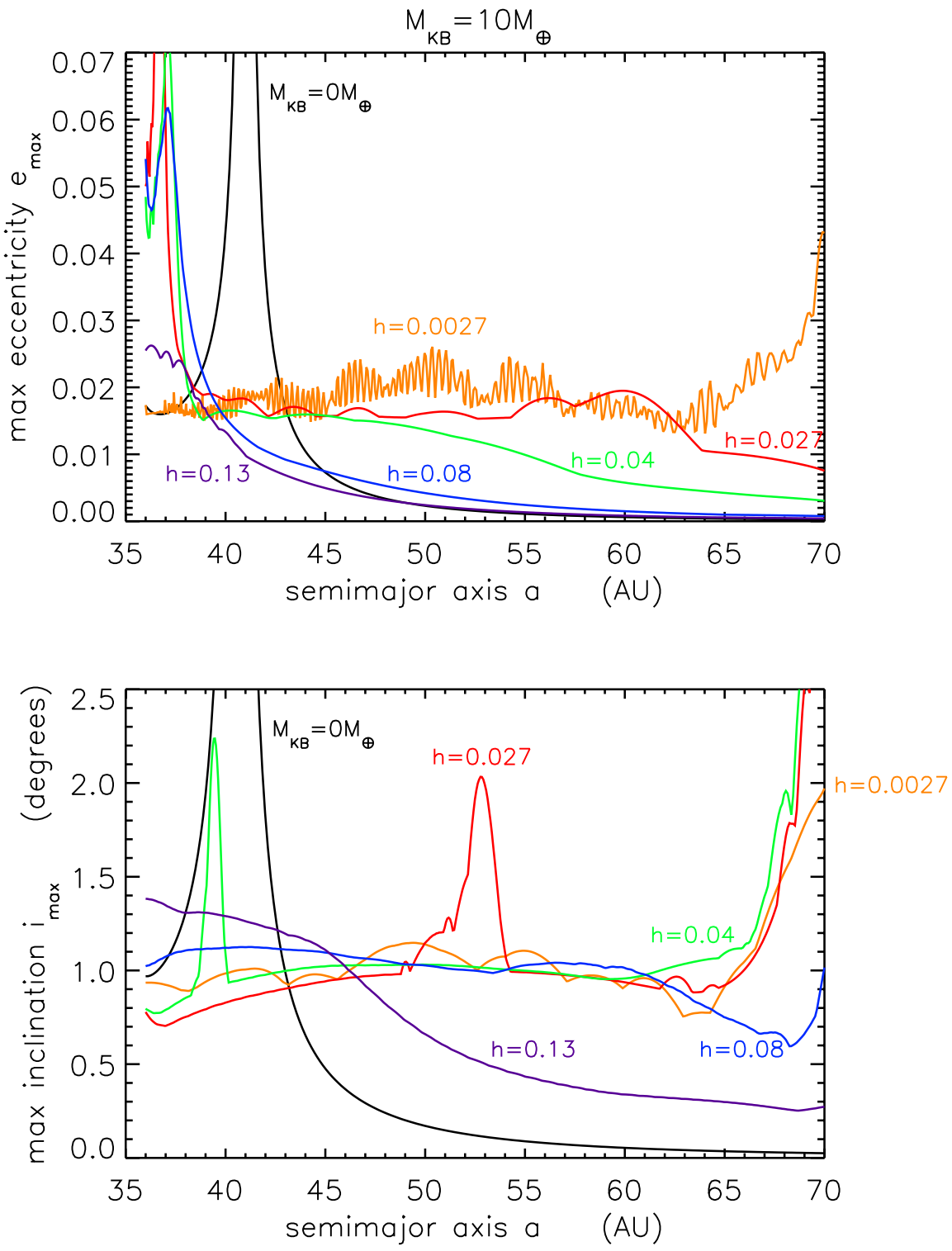

FIG. 5.—Maximum eccentricities and inclinations in a $M_{\rm KB} = 10\ M_{\oplus}$ disk having a fractional thickness $\mathfrak{h} = 0.0027$, 0.027, 0.04, 0. 08, and 0.13. The black curves are the forced e and i that occur in a massless disk. The characteristic particles size corresponding to each disk thickness $\mathfrak{h}$ can be obtained by setting particle dispersion velocities equal to their surface escape velocity, which corresponds to particle radii of $R \sim 5000\mathfrak{h}$ km $\sim$ 14, 140, 200, 400, and 650 km, respectively.

short waves completely dominate the disk's surface density structure.

Figure 1 also shows that the waves in lower mass systems have higher amplitudes. This suggests that wave action may tend to drive disk-planet systems toward an equipartition of angular momentum deficits, since the angular momentum content of the waves seen in all simulations is $\sim$0.5% and $\sim$10% of the planets' L_e and L_i, respectively.

The large-amplitude waves seen in the $M_{\rm KB} = 0.2\ M_{\oplus}$ disk also suggest that apsidal and nodal wave action might account for much of the Kuiper Belt's excited state (see Fig. 1). However, the viability of this scenario depends critically

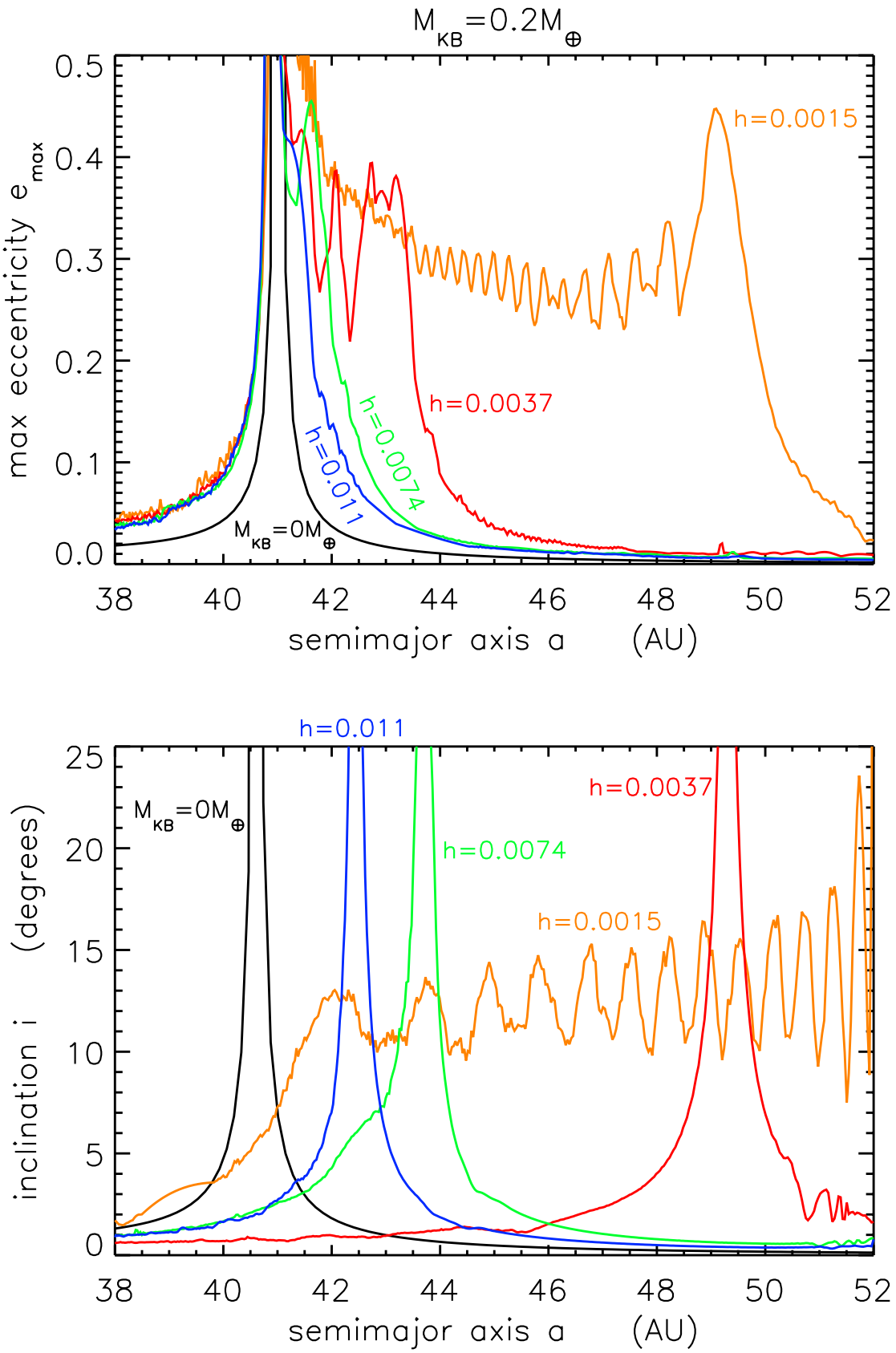

FIG. 6.—*Top*: Maximum eccentricities, $e_{\max}$, in a $M_{\rm KB} = 0.2\ M_\oplus$ disk having thicknesses $\mathfrak{h} = 0.0015$, 0.0037, 0.0074, and 0.011. *Bottom*: *Average* inclinations i that occur after the bending wave has stalled. The exception is the orange $\mathfrak{h} = 0.0015$ curve; this wave does not stall, so this curve shows the maximum inclination, $i_{\max}$. The black curves are the forced e and i that occur in a massless disk.

on the timescales that govern the belt's mass loss as well as the KBOs' velocity evolution. First note that accretion models show that as soon as the large $R \gtrsim 100$ km KBOs form, further KBO growth is halted as they initiate an episode of more vigorous collisions that steadily grind the belt's smaller members down to dust, which radiation forces then transport away (Kenyon & Luu 1999; Kenyon & Bromley 2001). Accretion models show that the $R \sim 100$ km KBOs form at $a \sim 35$ AU over a $\tau_{\rm form} \sim 1 \times 10^7$ yr timescale, and that the belt's subsequent collisional erosion occurs over a $\tau_{\rm erode} \sim 5 \times 10^8$ yr timescale (Kenyon & Luu 1999; Kenyon & Bromley 2001). If Neptune's orbit had migrated substantially, the advancing 2 : 1 resonance would also have swept across the Main Belt, which further enhances the stirring as well the collisional/dynamical erosion. The resulting grinding and erosion thus makes it ever more difficult for the belt to sustain waves as the initially massive Kuiper Belt mass is eroded by a factor of $\sim$150, which also lowers the disk's $\mathfrak{h}_Q \rightarrow \sim 10^{-3}$. Note that the removal of the smaller KBOs also turns off the dynamical friction that once kept the particle disk quite thin. The gravitational stirring by the surviving larger KBOs is then free to raise their dispersion velocities c up to their surface escape velocity $v_{\rm esc}$ (Safronov 1972), which results in a thicker disk with $\mathfrak{h} \sim v_{\rm esc}/na \sim 0.02(R/100\ {\rm km})$, where R is the KBOs' characteristic size. Since this is substantially larger than the current belt's $\mathfrak{h}_Q$, it seems quite likely that this stirring shut off the waves while they were still of low amplitude (see Fig. 1), long before the Kuiper Belt was ground down to its present mass.[6] Consequently, apsidal and nodal waves were likely able to propagate throughout the Kuiper Belt during a timescale $\tau_{\rm wave}$ that is bounded by the moment when the large KBOs first formed and when the belt eroded away, i.e., $\tau_{\rm form} < \tau_{\rm wave} < \tau_{\rm erode}$.

4.5. *Comments on Studies of Spiral Waves in Galactic Disks*

It should also be noted that this implementation of the rings model does not account for any viscous damping of spiral waves which, as Hunter & Toomre (1969) point out, may be more important as bending waves approach a disk's outer edge. Unlike the sharp-edged disks employed here, a more realistic disk likely has a "fuzzy" edge where the disk's surface density gently tapers to zero. Hunter & Toomre (1969) show that bending waves entering such a lower density zone tend to excite substantially larger inclinations there, and such sites will be considerably more susceptible to the viscous dissipation of these possibly nonlinear waves.

Toomre (1983) suggests that disks having a tapered edge might also cause bending waves to stall there, since the group velocity $c_g \rightarrow 0$ as the surface density σ smoothly goes to zero. However, this assertion was not confirmed by the rings model, which tapered a disk's outer surface density by multiplying it by the factor $[1 - (l - \Delta a)^2/l^2]^{1/2}$, where Δa is the distance from the outer edge and l is the tapering scale length. These experiments adopt values of l that are smaller than, comparable to, and larger than the bending wavelength λ_L, and in all cases the bending wave reflects at or near the disk edge, with considerably larger inclinations being excited in this tapered zone.

[6] An exception to this assertion might occur if the reflection of nodal waves at an outer edge allowed the belt to behave as a resonant cavity (Ward 2003). When the disk's mass and the wave's pattern speed are appropriately tuned, which can occur naturally as the belt eroded and/or as Neptune's precession rate varied due to its orbital migration, then a higher amplitude standing wave pattern can result while the belt persists in the tuned state.

5. SUMMARY

A model that rapidly computes the secular evolution of a gravitating disk-planet system is developed. The disk is treated as a nested set of rings, with the rings'/planets' time evolution being governed by the Lagrange planetary equations. It is shown that the solution to the dynamical equations is a modified version of the classical Laplace-Lagrange solution for the secular evolution of the planets (Brouwer & Clemence 1961; Murray & Dermott 1999), with the modification being due to a ring's finite thickness $h = c/n$ that is a consequence of the dispersion velocity c of that ring's constituent particles. Since the ring's finite thickness h softens its gravitational potential, this also softens the Laplace coefficients appearing in the Laplace-Lagrange solution over a scale h/a.

It is shown that the Lagrange planetary equations admit spiral wave solutions when the tight-winding approximation is applied. There are two types of spiral density (or apsidal) waves, long waves of wavelength $\lambda_L = 2\pi\mu_d|n/\omega|a$ and short waves of wavelength $\lambda_S < 2\sqrt{2}\pi\mathfrak{h}a$, where $\mathfrak{h} = h/a$ is the disk's fractional thickness and ω is the angular rate at which the spiral pattern rotates. The simulations presented here show that the giant planets launch long waves either at a resonance in the disk or at the disk's nearest edge, and that these waves propagate away until they reflect at the disk's far edge or else at a Q barrier in the disk, which resides where $\mathfrak{h} = \mathfrak{h}_Q(a)$, where $\mathfrak{h}_Q(a) = 0.26\mu_d|n/\omega|$ is the maximum disk thickness that can sustain apsidal waves, with $\mu_d = \pi\sigma a^2/M_\odot$ being the normalized disk mass. Of course, all of these findings can be derived from the stellar dispersion relation given in Toomre (1969) in the limit that the pattern speed ω is much smaller than the disk's mean motion n. Nonetheless, it is satisfying to see that the theory of unforced apsidal waves is readily obtained from the Lagrange planetary equations; with a little more effort the theory for forced apsidal waves (e.g., Ward & Hahn 1998a) should also be recoverable.

However, new results are obtained for the nodal wave problem, which admits only a long-wavelength solution λ_L to the planetary equations in the tight-winding limit. In particular, it is shown that these waves can stall, that is, the waves' group velocity plummets to zero as they approach a site in the disk where $\mathfrak{h} = 2.72\mathfrak{h}_Q$. If, however, these waves instead encounter a disk edge, they will reflect and return as long waves. In the limit that $\mathfrak{h} \rightarrow 0$, the results for nodal waves propagating in a infinitesimally thin disk is recovered (Ward & Hahn 2003), but note that the wave-stalling phenomenon does not appear in a $\mathfrak{h} = 0$ treatment of the disk.

The rings model is also used to examine the propagation of apsidal and nodal waves that are launched by the giant planets into a variety of Kuiper Belts having a mass $M_{\rm KB} = 30\ M_\oplus$ (the estimated primordial mass) down to $M_{\rm KB} = 0.08\ M_\oplus$ (which is $\sim$40% of the belt's current mass estimate). In each simulation the giant planets deposit roughly the same fraction of their initial angular momentum deficits, $\sim$0.5% and $\sim$10% of the planets' L_e and L_i, respectively, into the disk in the form of spiral waves. And

since the waves' angular momentum content is roughly the same in each simulation, the lower mass Kuiper Belts thus experience higher amplitude waves. Indeed, the waves seen in the $M_{\rm KB} \leq 0.2\ M_\oplus$ simulations are of sufficient amplitude that they could in principle account for much of the dynamical excitation that is observed in the Kuiper Belt. However, wave action in a $M_{\rm KB} \sim 0.2\ M_\oplus$ belt also requires its fractional scale height to be quite thin, namely, $\mathfrak{h} \lesssim 10^{-3}$. Most likely, apsidal and nodal waves were shut off, due to self-stirring by large KBOs as well as by other external perturbations, long before the belt eroded down to its current mass, in which case the excitation by wave action would have been quite modest.

The rings model developed here has many other applications. One issue of great interest is to determine whether apsidal and nodal waves may be propagating in Saturn's rings. Of particular interest are the short apsidal waves, since their detection could yield the ring's dispersion velocity c via a measurement of the short wavelength $\lambda_S \sim 9c/n$. Although the ring particles' dispersion velocity is of fundamental importance to ring dynamics, it is less than well constrained at Saturn. Of course, the differential precession due to planetary oblateness also needs to be included in the model (e.g., Murray & Dermott 1999), since this effect may actually defeat this form of wave action. The rings model can also be used to examine the forced motions of a relatively massless but much thicker circumstellar dust disk like β Pictoris. The warps and brightness asymmetries seen in this system are usually attributed to secular perturbations exerted by an unseen planetary system, and the code developed here can be used to very rapidly explore the wide range of planetary parameters. This rings model will be used to study these and other problems in greater detail in the near future.

This paper is contribution 1165 from the Lunar and Planetary Institute, which is operated by the Universities Space Research Association by cooperative agreement NCC5–679 with the National Aeronautics and Space Administration. This research was also supported by NASA via Origins of Solar Systems grant NAG5–10946, issued through the Office of Space Science.

APPENDIX A

Differentiating the $\tilde{b}_s^{(m)}$ appearing in the f and g functions (eqs. [17]) yields

$$f = 2\alpha\tilde{b}_{3/2}^{(1)} + \tfrac{3}{2}\alpha^2\left(\tilde{b}_{5/2}^{(2)} - \tilde{b}_{5/2}^{(0)}\right) - 3\alpha^2H^2(2+H^2)\tilde{b}_{5/2}^{(0)}\ , \tag{A1a}$$

$$g = 2(\alpha^2+1)(1+H^2)\tilde{b}_{3/2}^{(1)} - 3\alpha\left(\tilde{b}_{3/2}^{(0)} + \tilde{b}_{3/2}^{(2)}\right) + 3\alpha^2H^2(2+H^2)\tilde{b}_{5/2}^{(1)} - \tfrac{3}{4}\alpha^2\left(\tilde{b}_{5/2}^{(3)} - \tilde{b}_{5/2}^{(1)}\right)\ , \tag{A1b}$$

where $H^2 = \frac{1}{2}(\mathfrak{h}^2 + \mathfrak{h}'^2)$. The recursion relations

$$m\tilde{b}_s^{(m)} = s\alpha\left(\tilde{b}_{s+1}^{(m-1)} - \tilde{b}_{s+1}^{(m+1)}\right)\ , \tag{A2a}$$

$$(m+1-s)\alpha\tilde{b}_s^{(m+1)} = m(1+\alpha^2)(1+H^2)\tilde{b}_s^{(m)} - (m+s-1)\alpha\tilde{b}_s^{(m-1)}\ , \tag{A2b}$$

can be used to simplify equations (A1) further. Equations (A2) are derived in Brouwer & Clemence (1961) for the case where $H = 0$. However, the more general relations given above are readily obtained by replacing the combination $1+\alpha^2$ appearing in the Brouwer & Clemence (1961) recursion relations with $(1+\alpha^2)(1+H^2)$. So for $m = 1$ and $s = 3/2$, equation (A2a) is

$$\tilde{b}_{3/2}^{(1)} = \tfrac{3}{2}\alpha\left(\tilde{b}_{5/2}^{(0)} - \tilde{b}_{5/2}^{(2)}\right) \tag{A3}$$

and

$$\tilde{b}_{3/2}^{(2)} = \tfrac{3}{4}\alpha\left(\tilde{b}_{5/2}^{(1)} - \tilde{b}_{5/2}^{(3)}\right) \tag{A4}$$

for $m = 2$ and $s = 3/2$, while equation (A2b) yields

$$2(\alpha^2+1)(1+H^2)\tilde{b}_{3/2}^{(1)} = \alpha\tilde{b}_{3/2}^{(2)} + 3\alpha\tilde{b}_{3/2}^{(0)} \tag{A5}$$

for $m = 1$ and $s = 3/2$. Inserting equations (A3)–(A5) into (A1) then yields

$$f(\alpha,\mathfrak{h},\mathfrak{h}') = \alpha\tilde{b}_{3/2}^{(1)} - 3\alpha^2H^2(2+H^2)\tilde{b}_{5/2}^{(0)}\ , \tag{A6a}$$

$$g(\alpha,\mathfrak{h},\mathfrak{h}') = -\alpha\tilde{b}_{3/2}^{(2)} + 3\alpha^2H^2(2+H^2)\tilde{b}_{5/2}^{(1)}\ . \tag{A6b}$$

APPENDIX B

The symbolic mathematics software MAPLE has been used to write the needed softened Laplace coefficients $\tilde{b}_s^{(m)}$ (eq. [12]), in terms of complete elliptic integrals K and E. Setting $H^2 = \frac{1}{2}(\mathfrak{h}^2 + \mathfrak{h}'^2)$, $\gamma = (1+\alpha^2)(1+H^2)/2\alpha$, and

$\chi = [2/(\gamma+1)]^{1/2}$, then

$$\tilde{b}^{(0)}_{1/2}(\alpha,\mathfrak{h},\mathfrak{h}') = \frac{4K(\chi)}{\pi\sqrt{2\alpha(\gamma+1)}} \tag{B1a}$$

$$\tilde{b}^{(1)}_{1/2}(\alpha,\mathfrak{h},\mathfrak{h}') = \frac{4[\gamma K(\chi)-(\gamma+1)E(\chi)]}{\pi\sqrt{2\alpha(\gamma+1)}} \tag{B1b}$$

$$\tilde{b}^{(0)}_{3/2}(\alpha,\mathfrak{h},\mathfrak{h}') = \frac{2E(\chi)}{\pi\alpha(\gamma-1)\sqrt{2\alpha(\gamma+1)}} \tag{B1c}$$

$$\tilde{b}^{(1)}_{3/2}(\alpha,\mathfrak{h},\mathfrak{h}') = \frac{2[-(\gamma-1)K(\chi)+\gamma E(\chi)]}{\pi\alpha(\gamma-1)\sqrt{2\alpha(\gamma+1)}} \tag{B1d}$$

$$\tilde{b}^{(2)}_{3/2}(\alpha,\mathfrak{h},\mathfrak{h}') = \frac{2[-4\gamma(\gamma-1)K(\chi)+(4\gamma^2-3)E(\chi)]}{\pi\alpha(\gamma-1)\sqrt{2\alpha(\gamma+1)}} \tag{B1e}$$

$$\tilde{b}^{(0)}_{5/2}(\alpha,\mathfrak{h},\mathfrak{h}') = \frac{4[-(\gamma-1)K(\chi)+4\gamma E(\chi)]}{3\pi(2\alpha)^{5/2}(\gamma+1)^{3/2}(\gamma-1)^2} \tag{B1f}$$

$$\tilde{b}^{(1)}_{5/2}(\alpha,\mathfrak{h},\mathfrak{h}') = \frac{4[-\gamma(\gamma-1)K(\chi)+(\gamma^2+3)E(\chi)]}{3\pi(2\alpha)^{5/2}(\gamma+1)^{3/2}(\gamma-1)^2}\ , \tag{B1g}$$

where

$$K(\chi) = \int_0^1 \frac{dt}{\sqrt{(1-t^2)(1-\chi^2t^2}} \tag{B2}$$

is the complete elliptic integral of the first kind, and

$$E(\chi) = \int_0^1 \sqrt{\frac{1-\chi^2t^2}{1-t^2}}\,dt \tag{B3}$$

is the complete elliptic integral of the second kind. The series expansions for $K(\chi)$ and $E(\chi)$ given in Abramowitz & Stegun (1970) then permit the rapid calculation of the softened Laplace coefficients $\tilde{b}^{(m)}_s$ without requiring a numerical integration of equation (12). However, equation (B1) can give unreliable results for extreme values of α because of numerical roundoff errors. In this case it is preferable to factor the $(1+\alpha^2)(1+H^2) = 2\alpha\gamma$ term out of the integrand in equation (12) and expand the denominator for the case of large $\gamma \gg 1$:

$$\tilde{b}^{(m)}_s = \frac{2}{\pi(2\alpha\gamma)^s}\int_0^\pi d\phi\cos(m\phi)\left[1+\frac{s}{\gamma}\cos\phi+\frac{s(s+1)}{2\gamma^2}\cos 2\phi+\ldots\right] \tag{B4a}$$

$$\simeq \frac{f_m\alpha^m}{(2\alpha\gamma)^{s+m}}\ , \tag{B4b}$$

where $f_0 = 2, f_1 = 2s$, and $f_2 = s(s+1)$. Equation (B4b) usually gives the more reliable result for $\alpha \ll 0.01$ and for $\alpha \gg 100$.

Another useful form for $\tilde{b}^{(m)}_s$ is obtained for regions where $\alpha = 1+x$ where $|x| \ll 1$. In this "local" approximation, the dominant contribution to the integral (eq. [12]), occurs where $\phi \ll 1$. Thus, we can set $\cos\phi \simeq 1-\phi^2/2$ and $\alpha \simeq 1$ except where it appears as $\alpha - 1 = x$, extend the upper integration limit to infinity, and set $\cos(m\phi) \simeq 1$ in the numerator (e.g., Goldreich & Tremaine 1980):

$$\tilde{b}^{(m)}_s(x) \simeq \frac{2}{\pi}\int_0^\infty \frac{d\phi}{(x^2+2H^2+\phi^2)^s} = \begin{cases} \dfrac{2/\pi}{x^2+2H^2} & \text{for } s = 3/2\ , \\ \dfrac{4/3\pi}{(x^2+2H^2)^2} & \text{for } s = 5/2\ . \end{cases} \tag{B5}$$

APPENDIX C

The time derivative of L_e (eq. [26a]) is

$$\frac{dL_e}{dt} = \sum_j m_jn_ja_j^2\left(h_j\frac{dh_j}{dt}+k_j\frac{dk_j}{dt}\right) \tag{C1a}$$

$$= \sum_j\sum_{k\neq j} m_jn_j^2a_j^2A_{jk}(h_jk_k-h_kk_j) \tag{C1b}$$

$$= \sum_j\sum_{k\neq j}\frac{1}{4}\left(\frac{m_jm_k}{M_\odot+m_j}\right)n_j^2a_j^2g_{jk}(h_jk_k-h_kk_j)\ , \tag{C1c}$$

where $g_{jk} = g(\alpha_{jk}, \mathfrak{h}_j, \mathfrak{h}_k)$. Let $dL_e/dt = S_1 - S_2$, where S_1 is the sum over the $h_j k_k$ terms and S_2 is the sum over the $h_k k_j$. Swap the j and k indices in S_2 so that it becomes

$$S_2 = \sum_k \sum_{j \neq k} \frac{1}{4} \left(\frac{m_k m_j}{M_\odot + m_k} \right) n_k^2 a_k^2 g_{kj} h_j k_k \ . \tag{C2}$$

Equation (18c) shows that $g_{kj} = \alpha_{jk} g_{jk}$, and with $(n_k/n_j)^2 = \alpha_{jk}^{-3}(M_\odot + m_k)/(M_\odot + m_j)$, S_2 becomes

$$S_2 = \sum_k \sum_{j \neq k} \frac{1}{4} \left(\frac{m_j m_k}{M_\odot + m_j} \right) n_j^2 a_j^2 g_{jk} h_j k_k \ , \tag{C3}$$

which is S_1 since the sums obey $\sum_k \sum_{j \neq k} = \sum_j \sum_{k \neq j}$. Consequently, $dL_e/dt = S_1 - S_2 = 0$, and a similar analysis will also show that $dL_i/dt = 0$.

REFERENCES

Abramowitz, M., & Stegun, I. 1970, Handbook of Mathematical Functions (New York: Dover)
Borderies, N., Goldreich, P., & Tremaine, S. 1985, Icarus, 63, 406
Brouwer, D., & Clemence, G. 1961, Methods of Celestial Mechanics (New York: Academic Press)
Brown, M. E. 2001, AJ, 121, 2804
Chiang, E. I., & Jordan, A. B. 2002, AJ, 124, 3430
Duncan, M. J., & Levison, H. F. 1997, Science, 276, 1670
Fernandez, J. A., & Ip, W.-H. 1984, Icarus, 58, 109
Goldreich, P., & Tremaine, S. 1980, ApJ, 241, 425
Gomes, R. 2003, Icarus, 161, 404
Hahn, J. M., & Malhotra, R. 1999, AJ, 117, 3041
Hahn, J. M., & Ward, W. R. 2002, Lunar and Planetary Science Conference XXXIII, Abstract 1930 (CD ROM; Houston: Lunar and Planetary Inst.)
Hahn, J. M., Ward, W. R., & Rettig, T. W. 1995, Icarus, 117, 25
Heppenheimer, T. A. 1980, Icarus, 41, 76
Hunter, C., & Toomre, A. 1969, ApJ, 155, 747
Jacobs, V., & Sellwood, J. A. 2001, ApJ, 555, L25
Jewitt, D., Luu, J., & Trujillo, C. 1998, AJ, 115, 2125
Kenyon, S. J., & Bromley, B. C. 2001, AJ, 121, 538
Kenyon, S. J., & Luu, J. X. 1999, AJ, 118, 1101
Malhotra, R. 1995, AJ, 110, 420
Murray, C., & Dermott, S. 1999, Solar System Dynamics (Cambridge: Cambridge Univ. Press)
Nagasawa, M., & Ida, S. 2000, AJ, 120, 3311
Safronov, V. S. 1972, NASA TT F-677 (Washington: NASA)
Thommes, E. W., Duncan, M. J., & Levison, H. F. 2002, AJ, 123, 2862
Toomre, A. 1969, ApJ, 158, 899
———. 1983, in IAU Symp. 100, Internal Kinematics and Dynamics of Galaxies (Dordrecht: Reidel), 177
Tremaine, S. 2001, AJ, 121, 1776
Ward, W. R. 1981, Icarus, 47, 234
———. 2003, ApJ, 584, L39
Ward, W. R., & Hahn, J. M. 1998a, AJ, 116, 489
———. 1998b, Science, 280, 2104
———. 2003, AJ, 125, 3389