Annu. Rev. Astron. Astrophys. 1993. 31: 265–95

THE LONG-TERM DYNAMICAL EVOLUTION OF THE SOLAR SYSTEM

Martin J. Duncan

Department of Physics, Queen's University, Kingston, Ontario, Canada K7L 3N6 and Institute for Theoretical Physics, University of California, Santa Barbara, California 93106

Thomas Quinn

Department of Astrophysics, Oxford University, United Kingdom

KEY WORDS: chaos, stability

1. INTRODUCTION

The issue of the long-term stability of the Solar System is of course one of the oldest unsolved problems in Newtonian physics, but recent (largely numerical) work has provided some insight into the problem. In particular, several lines of investigation suggest that the Solar System is subject to the deterministic chaos recently found in many nonlinear Hamiltonian systems. This inherent unpredictability has profound implications for the dynamics of the Solar System, not the least of which is the demise of the "clockwork" picture which dominated thinking in the nineteenth century and carried over into much of the twentieth. However, the same numerical simulations that provide evidence for chaos also show that in most cases the timescale for macroscopic changes in the system (e.g. large changes in eccentricity, crossing of orbits, ejection of bodies, etc) is several orders of magnitude longer than the timescale for unpredictability. In practice, this means that although we probably will not be able to calculate the exact location of the Earth in its orbit 100 million years into the future (or the

0066–4146/93/0915–0265$02.00

past) there is a reassuringly large probability that it will be in a low-inclination, low-eccentricity orbit with semimajor axis of very nearly 1 AU. Nonetheless, on sufficiently long timescales rather dramatic orbital changes can occur and for many minor bodies these timescales are much shorter than the age of the Solar System. Furthermore, it is important to understand the origin of this behavior for the present Solar System (and others like it around other stars) and to understand its implications for the late stages of planet formation.

We do not review here our current understanding of the process of planet formation (see e.g. Lissauer 1993 in this volume) but will concentrate on the dynamical evolution of the system once the planets have acquired most of their material and settled into nearly circular, nearly coplanar orbits. We return at the end of this review to the results of simulations of the dynamical evolution of the planets themselves. However, many aspects of the evolution of the orbits of "test" bodies (such as comets and asteroids) under the perturbing influence of one or more planets (*a*) are of considerable intrinsic interest, (*b*) can be numerically more tractable, and (*c*) may offer considerable insight into the more general dynamical problem. There have also been several developments in the study of nonlinear dynamical systems that can help us to understand the newly emerging picture.

Thus, we begin with a brief review of those aspects of the dynamics of Hamiltonian systems relevant to the gravitational N-body problem, with an emphasis on recent developments which pertain to the long-term evolution of orbits. With this mathematical machinery as a basis, we then describe in Section 3 some examples of secular resonances in the inner Asteroid Belt, chaotic motion near mean-motion resonances with Jupiter, and the origins of the Kirkwood gaps in the Asteroid Belt. We also discuss the extent to which it is possible that the structure in the outer Asteroid Belt has been shaped by the gravitational influences of the giant planets (particularly Jupiter).

In Section 4 we discuss numerical simulations designed to test the dynamical stability of test particles between the giant planets and beyond Neptune. We discuss the chaotic nature of orbits in most of this region and describe the scattering of icy planetesimals from between the giant planets into what is now the Oort cloud. We also consider the evidence that a trans-Neptunian Kuiper belt of leftover planetesimals may be the source of short-period comets. In the penultimate section we return to the larger question of the long-term stability of the planets themselves and present the evidence that our own system (and others like it) are chaotic. We conclude with a summary of the results of the previous sections and briefly discuss promising areas for future research.

2. HAMILTONIAN DYNAMICS AND THE SOLAR SYSTEM

A Hamiltonian system is one in which the equations of motion for the coordinates and associated momenta can be derived from a Hamiltonian function in the way familiar from classical mechanics. Thus, insofar as the Sun, planets, and minor bodies can be approximated as point masses interacting solely via their mutual gravitational forces, the Solar System can be viewed as a nonlinear Hamiltonian system. It is therefore not surprising that many of the mathematical techniques used in the study of other nonlinear dynamical systems can be applied to the problem at hand. In the remainder of this section we will simply summarize a few key concepts which will be useful in understanding the numerical results to be presented in subsequent sections. Interested readers will find more information in the superb reviews by Hénon (1983) and Berry (1978) on Hamiltonian systems in general and by Wisdom (1987) on chaotic dynamics in the Solar System in particular. A brief discussion of the distinction between quasi-periodic and chaotic orbits, and the features of Poincaré surfaces of section and area-preserving mappings is also given in a previous review article (Duncan & Quinn 1992).

Each body in a system of N self-gravitating particles will in general have three degrees of freedom (e.g. three Cartesian coordinates) and with each degree of freedom one can associate a generalized coordinate and its conjugate momentum. The state of the system at any instant in time may then be represented as a point in a $6N$-dimensional phase space and the evolution of the system is then a trajectory (line) in this space which begins at some point determined by the initial spatial coordinates and momenta for each particle. In some cases, there exist one or more independent single-valued functions of the spatial coordinates and momenta which are conserved along each trajectory of the system. Each of these functions is called an integral of the motion. (For a single particle moving in a given potential, the most familiar examples are its energy if the potential is time-independent and/or the components of its angular momentum if the potential is spherically symmetric.) If there exists an independent integral for every degree of freedom, then the system is said to be integrable and Hamilton's equations can be used to show that each trajectory is then confined to a region that is topologically a torus of dimension equal to one half of the dimension of the phase space. These trajectories are called "quasi-periodic" and are the general case for an integrable Hamiltonian. For systems with two or more degrees of freedom, however, there is no guarantee that there will be an integral for every degree of freedom, and what is often found in practice is that for some initial conditions in a given potential the trajectories are quasi-periodic and for other initial conditions the trajectories are found to be irregular (or in current parlance "chaotic") and are not as confined in phase space.

There is a key feature of the irregular orbits which we will use here as a definition of "chaos": Two trajectories which begin arbitrarily close in phase space in a chaotic region will typically diverge exponentially in time. Ironically, the timescale for that divergence in a given chaotic region does not typically depend on the initial conditions! Thus, if one computes the distance in 6-dimensional phase space $d(t)$ between two particles having an initially small separation, it can be shown that for quasi-periodic orbits, $d(t)-d(t_0)$ grows as a power of time t (typically linearly), whereas for irregular orbits $d(t)$ grows exponentially as $d(t_0)e^{\Gamma(t-t_0)}$, where Γ is conventionally called the Lyapunov exponent and Γ^{-1} is called the Lyapunov timescale. Thus, when plotted as a function of time, the quantity $\gamma \equiv \ln[d(t)/d(0)]/t$ for chaotic trajectories eventually levels off at a value which is the Lyapunov value in that region whereas γ continues to decline for quasi-periodic orbits (see e.g. Benettin et al 1978). We shall show examples of this behavior in subsequent sections. From this definition of chaos, we see that chaotic orbits show such a sensitive dependence on initial conditions that the detailed long-term behavior of the orbits is lost within several Lyapunov timescales. Even a perturbation as small as 10^{-10} in the initial conditions will result in a 100% discrepancy in about 20 Lyapunov times.

The Solar System is an excellent example of a system that is sometimes called "nearly integrable" in the sense that to lowest order the planets move on independent Keplerian orbits around the Sun and such a system is integrable. However, the mutual planetary interactions clearly add terms to the zeroth order Hamiltonian. Most of classical perturbation theory assumed that all problems are integrable and their perturbation expansions typically pushed the nonintegrable parts of the Hamiltonian to successively higher orders in the perturbation parameters such as the ratio of planetary masses to the solar mass, the planetary eccentricities, and inclinations. However, Poincaré (1892) showed that these perturbation series are in general divergent and have validity only over finite time spans. We will return to this issue in Sections 3 and 5, but note here that this problem of nonintegrability for certain initial conditions can at least qualitatively be understood because of mathematical work which is now collectively known as the KAM theorem (see e.g. Hénon 1983 for a review and references). In brief, for nearly integrable Hamiltonians, chaotic regions do not appear randomly but are associated with trajectories in the original problem in which the ratios of characteristic frequencies of the original problem are sufficiently well approximated by rational numbers—i.e. near resonances.

The simplest of these resonances to visualize are so-called mean-motion resonances, in which the orbital periods of two bodies are commensurable, but others (for example the secular resonances discussed below) can be important as well. Numerical integration of orbits (beginning with the classic paper of Hénon & Heiles 1964) has generally confirmed this basic picture—i.e. quasi-periodic and chaotic orbits are inextricably intermixed in phase space in the same sense that rational numbers are inextricably mixed with irrationals along the real number line, although the extent and importance of the chaotic zones depends on the strength of the perturbations and on the location in phase space. We will see how these concepts can be applied to orbits in the Solar System in succeeding sections.

3. EVOLUTION OF STRUCTURE IN THE ASTEROID BELT

In this section we use the long-term evolution of test particles in the Asteroid Belt to illustrate several of the concepts discussed above. We thus adopt as a working hypothesis what has come to be called the "gravitational hypothesis"—namely that virtually all of the structure in the Asteroid Belt can be attributed to gravitational interactions with the planets which clear asteroids from certain regions over the age of the Solar System. This is supported by the coincidence of depletion and enhancement bands with mean motion resonances of Jupiter, the main gravitational perturber in the asteroid region, and with secular resonances to be discussed next. It is not yet clear whether the gravitational hypothesis can explain all of the details; in particular it may be that some of the depletions may be related to the late stages of planet formation (the "cosmogonic hypothesis"). This and other alternatives are reviewed in Greenberg & Scholl (1979).

Classically, much of the discussion of the long-term evolution of orbits in the Solar System used secular perturbation theory as its foundation. This is described in some detail in Chapter XVI of Brouwer & Clemence (1961) and we cannot reproduce the details here. Essentially the method involves writing the Hamiltonian as the sum of a part that describes the independent Keplerian motion of the planets about the Sun plus a part (called *the disturbing function*) that contains terms due to the pairwise interactions among the planets and the indirect terms associated with the back-reaction of the planets on the Sun. In general, one can then expand the disturbing function in terms of the small parameters of the problem (such as the ratio of planetary masses to the solar mass, the planetary eccentricities, and inclinations etc) as well as the other orbital elements of the planets, including the mean longitudes (i.e. the location of the planets in their orbits) and attempt to solve Hamilton's equations for the time-dependence of the planetary orbital elements. For long-term behavior, however, a fruitful approach (due to Lagrange and Laplace) involves averaging the disturbing function over the mean motions of the planets, resulting in what is known as the secular part of the disturbing function. If the disturbing function is further limited to terms of lowest order, the equations of motion of the orbital elements of the planets can be expressed as a coupled set of first-order linear differential equations. This system can then be diagonalized to find the proper modes, which are sinusoids, and the corresponding eigenfrequencies. The evolution of a given planet's orbital elements is, therefore, a sum of the proper modes. With the addition of higher order terms the equations are no longer linear; however, it is sometimes possible to find a solution of a form similar to the linear solution, except with shifted proper mode frequencies and terms involving combinations of the proper mode frequencies (Bretagnon 1974). We shall return to discuss the long-term validity of this method in Section 5, but now consider how this approach can give insight into asteroidal motion.

To lowest order in the planetary masses and eccentricities, the secular equations for a test particle are analogous to those of a driven harmonic oscillator which has a natural frequency equal to the precession frequency of the argument of perihelion which would be induced on the particle's orbit even by planets on circular orbits. The driving frequencies are the eigenfrequencies for the planetary eccentricity and inclination variations. It is well known that the solution for a driven oscillator is the sum of sinusoidal terms in which the amplitude of the term with the natural frequency determines the "free" or "proper" eccentricity in our context and in which the amplitude for a term with a given driving frequency has a divisor which is the difference between that frequency and the natural one. This is an example of what often occurs in expansions of the disturbing function and is called the problem of small divisors. Clearly as we move to a region where the two frequencies approach one another (a "secular resonance" in the current context) the amplitude of the forced oscillation formally diverges. The existence of secular resonances in the Asteroid Belt was noted as early as the 19th century (Le Verrier 1856), and it has long been evident that the inner edge of the Asteroid Belt near 2 AU is near such a resonance. Williams (1969, 1971) introduced a powerful approach to analyzing secular resonances, and showed that asteroids with $a < 2.6$ AU are severely depleted at the locations of the three strongest resonances. Further applications of this approach and numerical studies have accounted for much of the observed depletion in the inner regions of the Asteroid Belt (see Scholl et al 1989 for a recent review). Secular resonances may also play an important role in the stability of planetary orbits, as we shall see.

3.1 The Kirkwood Gaps

Let us now turn to what is perhaps more obvious structure in the Asteroid Belt—that associated with mean-motion resonances with Jupiter, in which a particle's period of revolution about the Sun is an integer ratio times Jupiter's period. An obvious mean-motion resonance is at the location of the Trojan asteroids—a 1:1 resonance with Jupiter. These asteroids librate about the points 60° behind or ahead of Jupiter and therefore never suffer a close approach to Jupiter. Another example of a protection mechanism provided by a resonance is the Hilda group at a mean motion 3:2 resonance. These asteroids have a libration about 0° of their critical argument, $\sigma \equiv 2\lambda' - \lambda - \bar{\omega}$, where λ' is Jupiter's longitude, λ is the asteroid's longitude, and $\bar{\omega}$ is the asteroid's longitude of perihelion (Schubart 1968). In this way, whenever the asteroid is in conjunction with Jupiter ($\lambda = \lambda'$), the asteroid is close to perihelion, ($\lambda' \approx \bar{\omega}$) and well away from Jupiter.

Using resonances to explain the gaps in the outer Asteroid Belt and the general depletion of the outer belt proves to be more difficult. A feature subject to much investigation has been the gap at the 3:1 mean motion resonance. Scholl & Froeschlé (1974) investigated this commensurability using the averaged planar elliptic restricted three-body problem. They found that most orbits starting at small eccentricity were regular and showed very little variation in eccentricity or semimajor axis over timescales of 50,000 yrs.

Investigation of more realistic models was limited because of the amount of computer time needed to follow orbits over long periods of time. A breakthrough occurred when Wisdom (1982) devised an algebraic mapping of phase space onto itself with a resonant structure similar to the 3:1 commensurability. His surprising result was that an orbit near the resonance could maintain a low ($e < 0.1$) eccentricity for nearly a million years and then have a sudden increase in eccentricity to over 0.3. Wisdom also showed that this behavior could not have been seen in the averaged planar eccentric restricted three-body problem because the averaged Hamiltonian admits a quasi-integral which confines the orbits to low eccentricities.

Any doubts about this apparent chaotic behavior were later dispelled by numerical integrations and the measurement of a nonzero maximum Lyapanov exponent (Wisdom 1983) and by a semianalytic perturbative theory which explained many of the features found in the mappings (Wisdom 1985). Figure 1 plots the eccentricity as a function of time for a typical chaotic trajectory near the 3:1 resonance as determined by direct integration. Note that the time is measured in Myr, so that the particle can remain in a low-eccentricity state for many tens of thousands of orbits before relatively rapidly entering a high-eccentricity phase. Approximate surfaces of section explaining this behavior may be found in Wisdom (1985). There it can be seen that the particle spends some time in low-eccentricity islands near the origin, while being free to explore a fairly large chaotic zone of higher eccentricity.

In Figure 2 we illustrate the distinction between adjacent trajectories that are regular versus those that are quasi-periodic as characterized by the Lyapunov exponent described in Section 2. The quantity $\gamma \equiv \ln [d(t)/d(0)]/t$ for chaotic trajectories eventually levels off at a value that is the inverse of the Lyapunov timescale for divergence, whereas γ continues to decline for quasi-periodic orbits when plotted as a function of time. The examples shown correspond to the orbits within and without a chaotic region of the Asteroid Belt near the 3:1 mean-motion resonance with Jupiter. In particular, the chaotic trajectory is the same as the one followed to compute eccentricity versus time in Figure 1. A comparison of the two Figures illustrates an important feature that often occurs in simulations (to be discussed later): The particle can remain in a low-eccentricity state for hundreds of Lyapunov times before "jumping" relatively quickly to high eccentricity. Wisdom (1987) gives other (albeit rarer) examples of

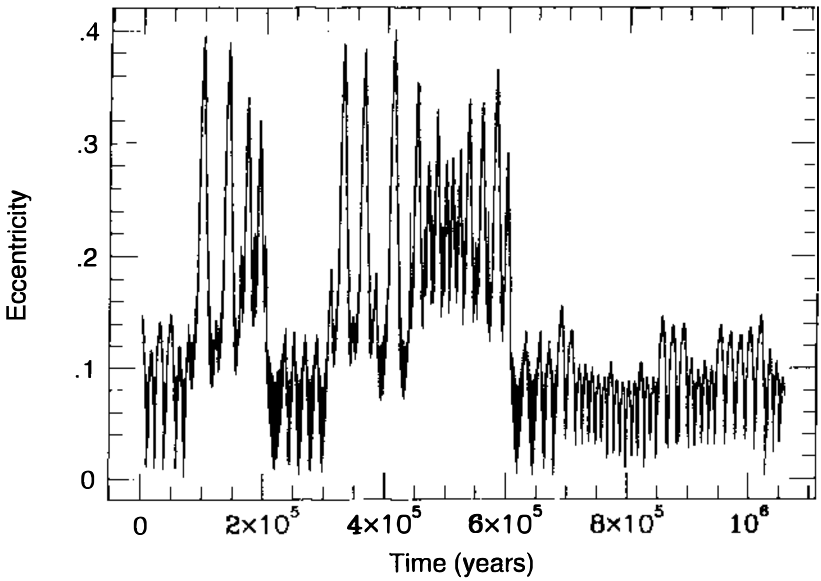

Figure 1 Eccentricity as a function of time for an integration of a chaotic trajectory near the 3:1 resonance in the elliptic restricted 3-body problem with parameters appropriate to the Sun-Jupiter-asteroid case.

trajectories that exhibit long periods of low-eccentricity punctuated by bursts of short-lived but highly eccentric behavior. The relevant surfaces of section clearly demonstrate that although a wide chaotic zone surrounds the origin, there also exists a narrow branch that extends to high eccentricity. An orbit wandering in the band near the origin could appear to be constrained to low eccentricities, and then suddenly suffer a large jump in eccentricity as it went down the branch.

The outer boundaries of the chaotic zone as determined by Wisdom's work has been shown to coincide well with the boundaries of the Kirkwood gap as shown in the numbered minor planets and the Palomar-Leiden survey. Since objects that begin on near-circular orbits in the gap acquire sufficient eccentricities to cross the orbit of Mars, the perturbative effects of Mars are believed capable of clearing out the 3:1 gap over the age of the Solar System. However, this has not actually been shown with direct integrations to date. Indeed, the study of orbits near resonances other than

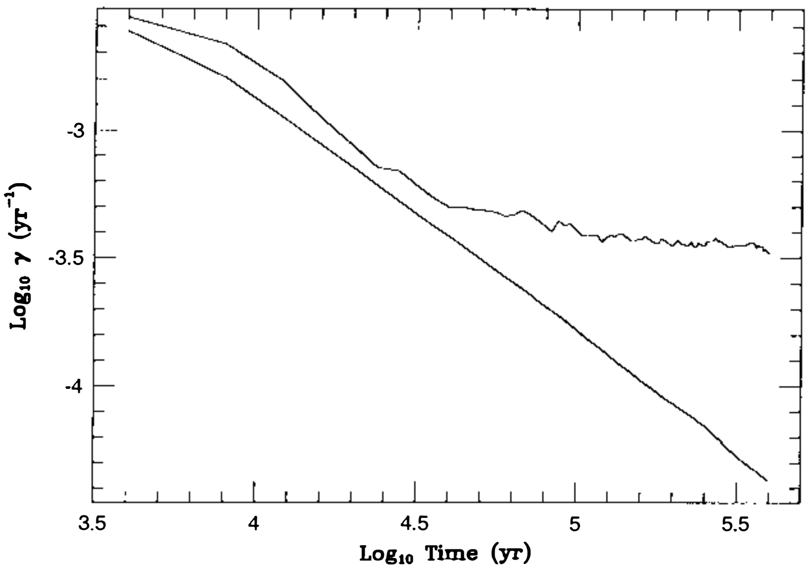

Figure 2 Distinction between regular (*lower curve*) and chaotic (*upper curve*) trajectories as characterized by the Lyapunov exponent discussed in the text. Both trajectories are near the 3:1 resonance in the elliptic restricted 3-body problem and the chaotic one is the same as that followed in Figure 1. For chaotic trajectories, a plot of log(γ) versus log(t) eventually levels off at a value of γ that is the inverse of the Lyapunov timescale for the divergence of initially adjacent trajectories. Note by comparison with Figure 1 that the chaotic particle can remain in a low-eccentricity state for hundreds of Lyapunov times before relatively rapidly entering a high-eccentricity phase.

the 3:1 remains an active area of research. In particular the very strong band near the 2:1 resonance as well as the other low-order resonances have been studied with a combination of quasi-analytic and numerical methods (see e.g. Lemaitre & Henrard 1990, Yoshikawa 1991, and references therein). These investigations typically find that regions near the resonances that are devoid of asteroids in the real Asteroid Belt show no obvious depletions in the simulations. It is possible that the simulations neglected subtle but cumulative effects or did not proceed long enough in the case of direct integrations, but it is also possible that cosmogonical effects have played a role in shaping some of the gaps. This is clearly an important issue which will be addressed in the next generation of simulations.

3.2 *The Outer Asteroid Belt*

The depletion of the Asteroid Belt exterior to the 2:1 resonance at 3.28 AU has long been assumed to be due to the gravitational influence of Jupiter (see e.g. Nobili 1989 for a review). This seems plausible because of the large number of strong resonances that occur in the outer belt. These resonances overlap at moderate eccentricities in regions that show a sharp cutoff in the asteroid distribution (Dermott & Murray 1983). The argument for depletion by Jupiter's perturbations was first checked by a numerical integration by Lecar & Franklin (1973) who numerically integrated orbits in the planar elliptic restricted three-body problem for a period of a few thousand years. They found that the region outside of 4.0 AU could be depleted in this time span, but the depletion inside 4.0 AU was very small. Froeschlé & Scholl (1979) extended this by integrating orbits in the 3-dimensional Sun-Jupiter-Saturn model for 10^5 yr. For small eccentricities, they again found very little depletion inside 4.0 AU. By carefully selecting their initial conditions, Milani & Nobili (1985) were able to find orbits that crossed Jupiter very quickly. However, for an initial semimajor axis less than 4.0 AU, only orbits with initially high eccentricities ($e > 0.15$) were found to be Jupiter-crossers in their relatively short integrations.

Gladman & Duncan (1990) performed 1.2×10^7 yr integrations of the equations of motion for the Sun, Jupiter, Saturn, and 80 test particles which began on circular orbits at various radii distributed between 3.1 and 3.9 AU. Particles which evolved into orbits that crossed that of Mars or which brought them to a close approach with Jupiter or Saturn were removed from the integrations. Gladman & Duncan found depletions at some of the stronger mean motion resonances, in particular the 2:1 resonance, but most of their orbits were stable over this timescale.

Lecar et al (1992) and Franklin & Lecar (1992) have extended their pioneering work on the original distribution of the asteroids by studying the orbits of several hundred objects in the outer Asteroid Belt for timescales of 10^7 to 10^8 years. Their results leave them "less optimistic for the gravitational hypothesis," based largely on an apparent correlation that they have found between the Lyapunov timescale of a given particle and the time for the particle to cross the orbit of a nearby planet. They argue that the crossing time is proportional to the Lyapunov timescale to the power 1.8, with a numerical coefficient that depends on the exact location in the Solar System. They use this result to suggest that, based on their computed Lyapunov times, most low-eccentricity asteroids in the region from just beyond the 2:1 resonance at 3.28 AU outward to 3.54 AU have inferred crossing times greater than 4.5 Gyr. Thus it appears likely that longer direct integrations will show considerable depletion over Gyr timescales in the region beyond the 2:1 resonance, but perhaps only in a population that begins on moderately eccentric orbits. If so, as emphasized by Dermott & Murray (1983), this may reflect the state of the Asteroid Belt at the end of the planet formation phase and would be an important constraint on models of planet formation. Accurate integrations on Gyr timescales in this region are computationally expensive but are required to settle this issue.

4. THE OUTER PLANETARY REGION AND COMETARY ORIGINS

Having discussed the dynamics in the Asteroid Belt, we turn now to a study of the evolution of test particles between and beyond the giant planets. We note in passing that the very long-term behavior of particles on low-eccentricity orbits between the terrestrial planets have not been thoroughly studied. The results of an approximate two-planet mapping approach (Duncan et al 1989) suggested that there may be a stable band between the Earth and Venus, but refinements of the map and numerical integrations suggest that this may not be the case (G. Quinlan, private communication). The work of Innanen and collaborators (see e.g. Mikkola & Innanen 1992, Innanen 1991, and references therein) show that the analog of the Trojan points at the 1:1 mean-motion commensurabilities of Venus, Earth, and Mars are stable locations, at least over their 2 Myr integration timescales. Indeed the asteroid 1990MB is the first known "Martian Trojan." Longer integrations of this entire terrestrial region would be of interest.

4.1 Chaotic Orbits Between the Giant Planets

In a sense, the Solar System seems to be remarkably devoid of objects. In the generally favored model for the formation of the Solar System, the planets accumulate through the accretion of vast numbers of planetesimals which formed on nearly circular orbits in the early Solar System. Since it is unlikely that all of these objects would have been incorporated into the present planets, a question that immediately arises is: Can any of these objects have survived to the present day? Because of the physical complications present in the early Solar System (e.g. gas drag, nebular dispersal, etc) one can ask the following simplified question: Are there any regions in the *current* Solar System where test particles (of negligible mass) placed on initially circular orbits are stable against having a future close approach with one of the giant planets? Having a close approach is not in itself a guarantee of ejection (although over sufficiently long timescales it usually is), but it usually means that the orbit is chaotic.

A complete answer to this question would of course require the direct integration over Gyr timescales of all of the planets together with many test particles distributed between them. This is not yet technically feasible, although it will be in the near future. On the analytical side, Wisdom (1980) showed that even for the planar, circular restricted three-body case there exists a band in semimajor axis centered on each planet's semimajor axis within which virtually all test particle trajectories are chaotic due to the overlap of first-order mean-motion resonances. Numerical investigations have shown that most test particles placed in these regions relatively quickly undergo a close encounter with the relevant planet. Test particle orbits between Jupiter and Saturn were numerically investigated on timescales of up to several Myr by Lecar & Franklin (1973) and Franklin et al (1989). In the first of these papers a band of semimajor axis was discovered in which initially low-eccentricity orbits were found to be stable over the timescales studied (500 Jupiter orbits). In the second paper the stable region (between 1.30 and 1.55 Jupiter semimajor axes) was studied more intensely, and the authors concluded that essentially no particles between Jupiter and Saturn were stable against becoming planet-crossers for longer than 10^7 years.

Gladman & Duncan (1990) have performed the most accurate integrations to date of the evolution of a swarm of test particles on initially circular orbits ranging from the outer Asteroid Belt to the Kuiper belt. The orbits of roughly one thousand test particles were followed for up to 22 Myr under the gravitational influence of the Sun and four (or in some cases two) of the giant planets. The mutual gravitational interactions of the planets were included. Test particles that underwent a close approach to a planet were removed from the integration.

The dynamical clearing of gaps found near some of the resonances with Jupiter in the Asteroid Belt has been discussed in Section 3. Exterior to Neptune there appeared to be some dynamical erosion of the inner edge of the Kuiper belt. Longer (but somewhat more approximate) integrations of the trans-Neptunian region are discussed in Section 4.4. The majority of the test particles *between* the giant planets were perturbed to a close approach to a planet on timescales of millions of years. Longer integrations of the same region using a modification of the Wisdom-Holman (1991) method (see Section 6) have been performed by Levison & Duncan (1993). The main modification was to calculate the motion of the giant planets from a synthetic secular perturbation theory using the dominant secular frequencies of Applegate et al (1986). The results agreed very well with those of Gladman and Duncan over the common range and extended them to 1 Gyr. The removal times as a function of initial semimajor axis for initially near-circular, low-inclination orbits between the giant planets are shown in Figure 3. Essentially all of the orbits become planet-crossing, often on surprisingly short timescales. This may explain the apparent absence of a large number of minor bodies between the giant planets but at first glance raises some rather disturbing questions about the stability of the planets themselves (see the next section) and may have implications in the field of planetary formation (see Gladman & Duncan 1990). In the next section we consider the evolution of those planetesimals initially between the giant planets that are gravitationally scattered to much larger semimajor axes.

4.2 *Evolution of the Oort Cloud*

The current version of the original theory of Oort (1950) for the formation of the cloud that bears his name is that comets formed as icy planetesimals on nearly circular orbits in the outer planetary system. Some of these accreted to form Uranus and Neptune, while the remainder were repeatedly scattered by the growing planets until they reached semimajor axes large enough for Galactic tidal fields and stellar perturbations to remove their perihelia from the planetary region, after which they are relatively immune to planetary perturbations. The bodies that reach "safety" in this way comprise the comet cloud which is present today. Tidal torquing by the smooth distribution in the Galactic disk has only recently been recognized as being more important (by a factor of 2–3) than stellar perturbations (Heisler & Tremaine 1986, Morris & Muller 1986, Smoluchowski & Torbett 1984, Torbett 1986) and may be responsible for a small nonrandomness in the distribution of observed long-period comets (Matese & Whitman 1989). Stellar perturbations are still important in randomizing the orbits in the cloud (see e.g. Weissman 1990) and occasional close stellar encounters can produce an unusually large influx of comets called comet showers (Hills 1981, Heisler 1990). Several competing theories of comet formation have been proposed (for example, that comets form in situ in the outskirts of an extended solar nebula), but they will

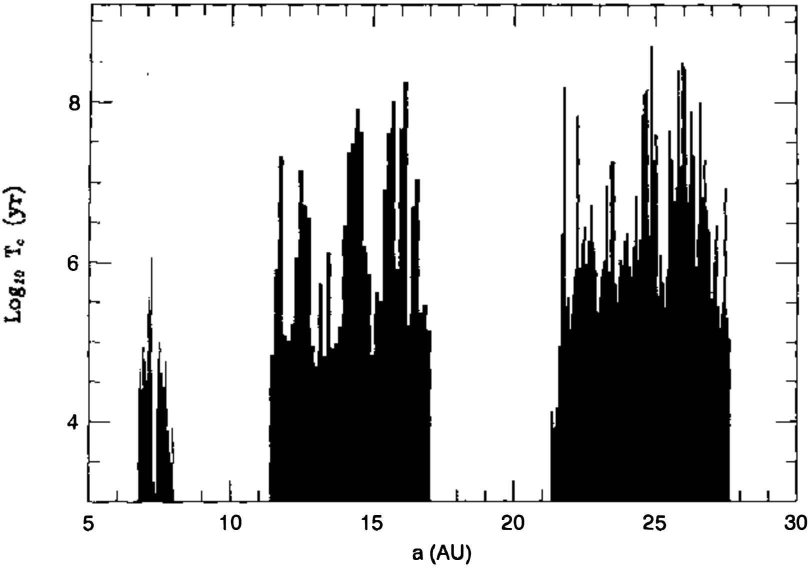

Figure 3 Diagram of removal times as a function of initial semimajor axis for test particles placed on low eccentricity, nearly circular orbits in the three bands between the giant planets. No particles were placed in the bands centered on each planet which were known to be unstable even in the single-planet case. (See text). The figure is similar to one in Gladman & Duncan (1990), but is obtained using the method described in Levison & Duncan (1993).

not be considered here (see Mumma et al 1992 and Bailey 1991 for reviews).

It has not yet been possible, of course, to follow the detailed evolution of the planetesimal swarm as the outer planets grew to their final masses. However, Duncan et al (1987) have picked up the story at the stage when the outer planets have attained their final masses and semimajor axes and have gravitationally scattered a large fraction of the unaccreted planetesimals to semimajor axes of a few hundred AU (although the perihelia of the comets remained in the planetary region). Details of the approach used and approximations made in the numerical simulations may be found in the original reference. (See also a simulation using an Öpik approximation by Shoemaker & Wolfe 1984). The simulations showed that planetary perturbations acting on moderately eccentric orbits caused a form of random walk in energy with little change in pericentric distance. Thus the formation of the comet cloud was driven by the interaction between planetary perturbations which drove diffusion in semimajor axis a at constant pericentric distance q, and Galactic tidal torques which changed q at fixed a, thereby removing cometary perihelia from the planetary region for sufficiently large a. Consequently, a typical comet evolved more or less in the ecliptic plane with pericenter near its birthplace until it was either ejected or torqued by the Galactic tidal field into a roughly spherical cloud with a more nearly isotropic velocity distribution. An inner edge to the cloud was found at ≈ 3000 AU—roughly the radius at which the timescales for the two effects are equal for comets formed in the Uranus-Neptune region. The density profile between 3000 and 50,000 AU is roughly a power law proportional to $r^{-3.5}$. The inner cloud ($a < 2 \times 10^4$ AU) thus contains roughly five times more mass than the classical Oort cloud.

Figure 4 depicts the evolution of the simulated comet swarm as a function of time. Note that the Galactic plane in each snapshot is a horizontal line, so that the ecliptic is inclined at an angle of $\approx 60°$ in these diagrams. The dotted circle in each snapshot is at a radius of 20,000 AU indicating the inner edge of the classical Oort cloud. It is evident from Figure 4 that the distribution after 10^8 yr was still biased toward the ecliptic for orbits with $a \lesssim 10^4$ AU. However, by 10^9 years the distribution was isotropic for $a \gtrsim 2000$ AU, and roughly 20% of the survivors were in the Oort cloud, with the remainder populating the inner cloud (except for the $\approx 4\%$ that became SP comets).

Between 10^9 and 4.5×10^9 yr, the total number of survivors decreased by a factor of ≈ 2, but the relative populations of the two clouds remained essentially unchanged. It is important to note, however, that the simulations did not include encounters with passing molecular clouds which may, over Gyr timescales, have a perturbing influence on the comets comparable to that of stars (Biermann 1978; Napier & Clube 1979; Bailey 1983, 1986a, 1991). This is particularly true for comets with $a \gtrsim 2.5 \times 10^4$ AU which are not protected by adiabatic invariance (i.e. those with orbital periods longer than the duration of a typical encounter). However, the influence of molecular clouds is difficult to estimate because their physical parameters are so uncertain. Indeed, the large observed fractions of wide binary stars with separations of order 20,000 AU suggests that the Oort cloud has probably not been disrupted by giant molecular clouds (Hut & Tremaine 1985, but see Weinberg et al 1986, 1987). Furthermore, the same events which strip comets from the classical Oort cloud will replenish that region with comets from the inner cloud. Nonetheless, a complete treatment of the long-term dynamical evolution of the Oort cloud will probably require the inclusion of molecular clouds (cf Heisler 1990).

4.3 *Origin of Short-Period Comets*

Having discussed the formation and evolution of the reservoir that supplies the current population of long-period comets, let us now turn to the origins of those with periods less than 200 years: the short-period (SP) population.

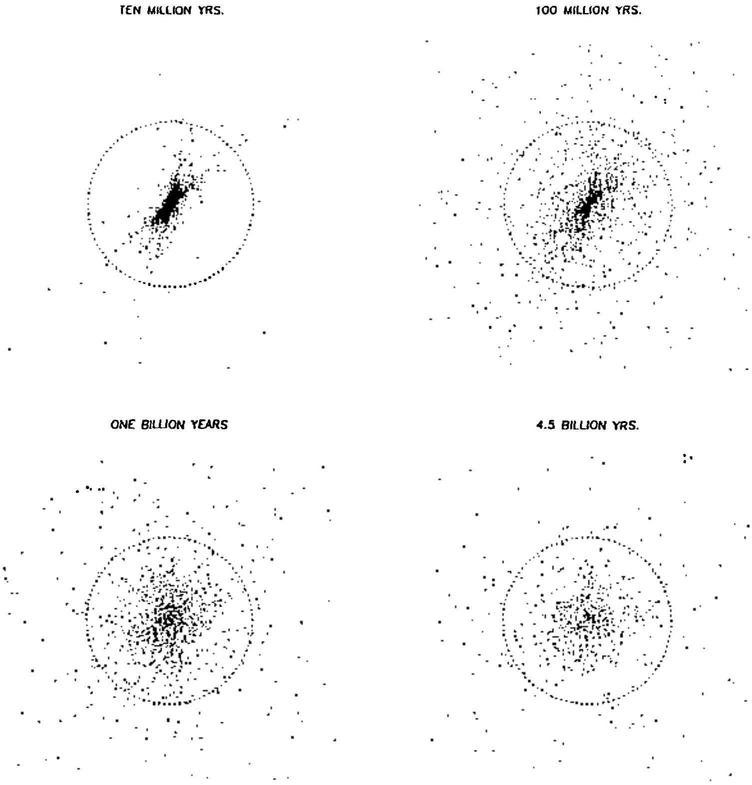

Figure 4 The formation of the Oort comet cloud in the simulation of Duncan et al (1987). The Galactic plane in each snapshot is a horizontal line. The dotted circle denotes a radius of 20,000 AU, indicating the inner edge of the classical Oort Cloud.

There are several striking features in the distribution of orbital elements of SP comets: 1. The distribution is strongly peaked towards periods $\lesssim 15$ yr. Roughly 80% of known SP comets have periods less than 15 years. 2. The SP comets are mostly on low-inclination prograde orbits: Only about 3% are on retrograde orbits and the inclinations i satisfy $\langle \cos i \rangle = 0.9$. (Recall that $\langle \cos i \rangle = 0$ for isotropic orbits and 1 for prograde orbits in the ecliptic.) 3. The arguments of perihelion ω (the angle between perihelion and ascending node) of the SP comets are strongly peaked near 0 and 180°.

Until recently, it has been believed that the SP comets originate in the Oort comet cloud. In an influential paper, Everhart (1972) argued from an extensive set of orbital integrations that repeated interactions with

Jupiter can produce SP orbits from near-parabolic orbits, so long as the initial inclination is small and the initial perihelion distance is near the orbit of Jupiter. He showed that the distribution of orbital elements for SP comets formed by this mechanism agreed well with observations. There was immediate concern, however, that Everhart's mechanism might not produce the correct number of SP comets. Joss (1973) has argued that the efficiency of this process is too low (by a factor of order 10^3–10^4) to produce the observed number of SP comets from the known flux of near-parabolic comets, although Delsemme (1973) argued that the discrepancy could be removed by the proper inclusion of all return passages of comets initially from the Oort cloud. (See Bailey 1992 and Yabushita & Tsujii 1991 for further references.)

In an ingenious Monte Carlo simulation, Everhart (1977) subsequently showed that a fraction of near-parabolic orbits with perihelion as large as Neptune's orbit will be gravitationally scattered by the outer planets into orbits that are Jupiter-crossing and that a fraction of these will eventually become SP comets. In conventional models of the Oort cloud, the efficiency of this process is too low to remove the flux discrepancy noted by Joss. However, Bailey (1986b) first suggested that if a very massive inner Oort cloud is present, the flux of near-parabolic comets into Neptune-crossing orbits may be sufficient to supply the SP comets. Subsequent work summarized in Bailey (1992) suggested that an inner cloud with a mass distribution similar to that described in Section 4.2 would suffice, but this source may overproduce high-inclination SP comets (see below).

An alternative theory proposes that SP comets originate in a belt of low-inclination comets just beyond the orbit of Neptune, between about 35 and 50 AU (e.g. Fernández 1980, Fernández & Ip 1983). The belt could be a natural remnant of the outermost parts of the solar nebula (Kuiper 1951, Whipple 1964), possibly producing a component of the infrared background at 100 μm detected by *IRAS* (Low et al 1984). *IRAS* data also indicate the presence of flattened dust shells around other stars, extending from 20 to 100 AU (see Weissman 1986 for a review). If some of the belt comets can be perturbed into Neptune-crossing orbits, subsequent scattering by the outer planets converts some of these into observable SP comets, in the manner described by Everhart (1977).

In order to test these two hypotheses, Duncan et al (1988: hereafter called DQT88) performed an extensive series of numerical integrations of a representative sample of comet orbits in the field of the Sun and the giant planets. Comets which began with perihelia either near Jupiter (q between 4 and 6 AU) or in the outer planetary region (q between 20 and 30 AU) were integrated until the comet was either ejected or became visible to an Earth-based observer, which the authors assumed occurred when $q < 1.5$ AU. A numerical obstacle that the authors had to overcome was imposed by the slow process of orbital evolution: Evolution from a Neptune-crossing orbit to a visible orbit typically takes millions of orbits. However, since gravitational scattering is a diffusion process, the authors argued that multiplying the mass of all the giant planets by a fixed factor μ should change the rate of evolution but not the statistical properties of the final distribution of orbits. (We shall return to this point below.)

The results showed that the inclination distribution of comets with large perihelion ($q \lesssim 30$ AU) that evolve to observable comets (i.e. those with $q \lesssim 1.5$ AU) is approximately preserved. Thus, the short-period comets, which are mostly in prograde, low-inclination orbits, cannot arise from gravitational scattering of any spherical population of comets (such as the Oort cloud). However, the distribution of orbital elements of SP comets arising from a population of low-inclination Neptune-crossing comets is in excellent agreement with observations. The authors concluded that the SP comets arise from a cometary belt in the outer solar system (now called the Kuiper belt).

Stagg & Bailey (1989) and Bailey & Stagg(1990) have argued that an inner Oort cloud may still be a viable source for the SP comets if the mass enhancement factor used by DQT88 badly underestimated the effects of close encounters and/or if there exists a large population of unobserved extinct high-inclination comets. Quinn et al (1990) have replied with a more extensive series of experiments, taking into account a variety of nongravitational and selection effects, and conclude that those SP comets with periods less than 20 yr (sometimes called the Jupiter family) cannot arise from an isotropic distribution. They argue that the results from simulations with $\mu = 40$ do not differ substantially from those with $\mu = 10$. Furthermore, Wetherill (1990) used an Öpik (1951) approximation using Arnold's (1965) method and $\mu = 1$ and obtained results very similar to DQT88. Bailey (1992) has pointed out that the Öpik approach seems to overestimate the capture probability by an order of magnitude relative to direct integrations, and so may also be suspect. A definitive resolution to the problem requires an extremely CPU-intensive direct integration with $\mu = 1$, with modeling for observational selection effects and the finite lifetimes of comets.

If the proposed Kuiper belt indeed exists, two questions naturally arise. 1. Since comets on Neptune-crossing orbits typically are ejected or evolve to SP comets on timescales that are much shorter than the age of the Solar System, what mechanism injects comets from the Kuiper belt into planet-crossing orbits? 2. What is the current structure of this belt and can one detect its larger members? We turn to these issues next.

4.4 *Structure in the Kuiper Belt*

It is possible that the minor body Chiron (Kowal 1979) originated in the Kuiper belt. Chiron, which is a roughly 100-km-sized object, is on a Saturn-crossing orbit which is unstable on a timescale of 10^5 to 10^6 years (Oikawa & Everhart 1979). It has recently been shown to exhibit cometary behavior (Luu & Jewitt 1990) such as the formation of a resolved coma (Meech & Belton 1990). The recently discovered minor body Pholus (1992AD) may be another member of the same class of large bodies in the outer Solar System. Because of the short lifetime of their current orbits, it seems likely that they are representative of a much larger population of similar objects which currently reside in the Kuiper belt. [Indeed Stern (1991) has argued that there must at one time have been a substantial number of 1000 km objects in the outer Solar System, although most by now must have been scattered beyond 100 AU.] Objects in the belt might be detected by their retrograde motion as seen at opposition from the Earth, which amounts to roughly $150/r$ arcseconds per hour, where r is the heliocentric distance of the object in AU. However, since they are detected optically by means of scattered sunlight, their brightness is proportional to r^{-4}. Thus, Chiron itself would be fainter than V = 22 if it were beyond 40 AU.

There have been four major proper motion surveys designed to search for objects in the outer Solar System, and with the exception of Chiron's discovery, the surveys have been fruitless. These surveys are reviewed in Levison & Duncan (1990), and it is noted there that the lack of detections cannot be used to place severe constraints on the mass of the Kuiper belt because the size distribution of the belt members is unknown. Indeed, if all of the proto-comets had sizes less than 40 km, then no objects beyond 25 AU would have been detectable in Levison & Duncan's survey. Furthermore, we have seen that there are now theoretical reasons to believe that the region from Jupiter's orbit to somewhere beyond Neptune's orbit has been gravitationally swept clean of planetesimals. If the inner edge of the putative Kuiper belt is beyond 40 AU, the detection of even Chiron-sized objects is difficult (but not impossible). Indeed, as this review was being completed, *IAU Circular No. 5611* announced the discovery of minor body 1992 QB1 by Luu & Jewitt (1993), which appears to be a Chiron-like body at a current distance of approximately 41 AU. Its orbital elements ($a = 44.4$ AU, $e = 0.11$, $i = 2.2°$) are consistent with it being the first of a large number of such objects to be discovered in the Kuiper belt.

Now let us consider the injection of planetesimals from the Kuiper belt into Neptune-crossing orbits and the resulting structure of the belt several Gyrs after its presumed formation. An approximate mapping technique described in Duncan et al (1989) suggested that most near-circular orbits with semimajor axes greater than 33 AU (i.e. in the proposed Kuiper belt) were stable against becoming Neptune-crossers over the lifetime of the Solar System. However, Torbett (1989) performed direct numerical integration of test particles in this region including the perturbative effects of the four giant planets, although the latter were taken to be on fixed Keplerian orbits. He found evidence for chaotic motion with an inverse Lyapunov exponent on the order of Myrs for moderately eccentric, moderately inclined orbits with perihelia between 30 and 45 AU (a "scattered disk"). Torbett & Smoluchowski (1990) extended this work and suggested that even particles with initial eccentricities as low as 0.02 are typically on chaotic trajectories if their semimajor axes are less than 45 AU. Except in a few cases, however, the authors were unable to follow the orbits long enough to establish whether or not most chaotic trajectories in this group led to Neptune-crossing.

Levison (1991) treated the problem as a Markov chain but his calculation of diffusion coefficients apparently mixed long-term sinusoidal variations in the orbital elements with secular drifts and the method overestimated the amount of orbital evolution. Levison & Duncan (1993) have recently studied the problem using the Wisdom-Holman method described in Section 5.1, modified in the manner described in Section 4.1, in which the orbital elements of the planets are obtained from secular perturbation theory. By comparison with more exact simulations, it appeared adequate to include only the direct gravitational influences of Uranus and Neptune, thereby allowing the integration of thousands of particles for up to 1 Gyr (and in a few cases 4.5 Gyr). Simulations were performed with initially low-inclination orbits with small to moderate initial eccentricities ($e < 0.2$) and each particle was integrated for 1 Gyr unless it was removed because it either crossed Neptune's orbit or came within Neptune's Hill sphere. The behavior of particles that were ultimately removed sometimes showed a secular drift in eccentricity, but often was reminiscent of that described earlier for objects in the 3:1 Kirkwood gap and between the giant planets—i. e. long periods of relatively low-eccentricity oscillations punctuated by a very rapid jump to Neptune-crossing eccentricity. However, in this case some of the jumps can occur on Gyr timescales, and this may provide the necessary supply of Neptune-crossers at the current epoch. Furthermore, the simulations show that although many of the orbits inward of 45 AU are chaotic, none of the particles in the study with Lyapunov timescales greater than about 1 Myr actually became a Neptune-crosser in 4.5 Gyr. Indeed there seemed to be a rough correlation between Lyapunov and removal times which was consistent with the results of Franklin & Lecar (1992) when appropriately scaled (see Section 3.2).

In Figure 5 we show the removal times as a function of initial semimajor axis for test particles beginning with eccentricities = 0.01. A dashed line is used if the particle survived until the end of the simulation (which was 1 Gyr in this case). Note the intricate structure of the distribution inside 41 AU, with essentially complete erosion inside 34 AU, followed by bands of differing longevity including an unusually unstable region just inside 36 AU. Some of these features persist for orbits of somewhat higher eccentricity, although most of the region interior to 42 AU is substantially eroded for eccentricities larger than 0.1. It is tempting to associate particularly unstable (or, conversely, longer-lived) regions with mean-motion or secular resonances with the planets, the latter as calculated for example by Heppenheimer (1979) or Knezevik et al (1991), but no clearcut case can be made at the moment. Direct integrations for 4.5 Gyr, which explicitly calculate the planetary orbits, are now feasible and should provide a better picture of this region. However, the current structure must to some extent reflect the currently ill-understood processes of gravitational scattering and physical collisions among icy planetesimals that led to the formation of the outer planets in the first place.

5. LONG-TERM STABILITY OF PLANETARY ORBITS

5.1 *Our Solar System*

We return now to one of the oldest problems in dynamical astronomy: whether the planets will continue indefinitely in nearly circular, nearly coplanar orbits. As we have described in Section 3, if a Hamiltonian system has an integral of motion for each degree of freedom, then the system will be quasi-periodic, as can be shown by expressing the Hamiltonian in terms of action-angle variables. The orbits will be confined to a multidimensional torus and the orbital elements should be describable by a sum of periodic terms in the sense that the Fourier transform of the time evolution of any planet's coordinates will involve only integer combinations of the fundamental frequencies (one per degree of freedom). Arnold (1961) has shown that such stable orbits would describe the Solar System if the masses, eccentricities, and inclinations were sufficiently small. The real Solar System, however, does not satisfy Arnold's requirements, so the question of its stability is unresolved. Laplace and Lagrange showed that, if the mutual planetary perturbations were calculated to first order in the masses, inclinations, and eccentricities, the orbits could indeed be described by a sum of periodic terms, indicating stability. Successive work by Brouwer & van Woerkom (1950), Bretagnon (1974), and others has shown that this is still the case if the perturbations are expanded to higher

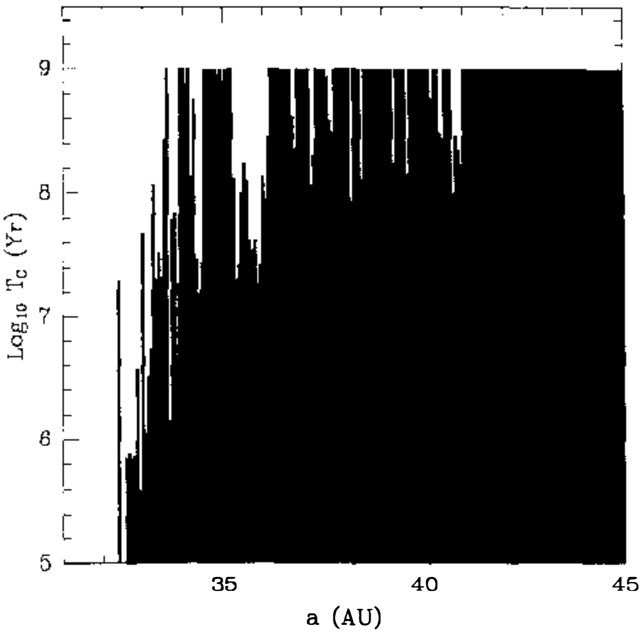

Figure 5 Diagram of removal times as a function of initial semimajor axis for test particles placed on low-inclination, nearly-circular orbits in the proposed Kuiper belt beyond Neptune. The method used to compute the planetary perturbations is described in the text. Particles were removed if they crossed Neptune's orbit or suffered a close approach to Neptune. Particles which survived until the end of the simulation (1 billion years) are also plotted. Note the resulting structure in the belt interior to 40 AU. The figure is based on the results in Levison & Duncan (1993).

orders. However, the work of Poincaré (1892) casts doubt on the long-term convergence of the various perturbation schemes. The problem with the perturbation expansion is that although the expansion is done in powers of small parameters, the existence of resonances between the planets will introduce small divisors into the expansion terms. Such small divisors can make high order terms in the power series unexpectedly large and destroy the convergence of the series.

We have briefly discussed in Section 3 the methods of secular perturbation theory and the problems associated with resonances. There are two separate points in the construction of the secular system at which resonances can cause nonconvergence of the expansion. The first is in averaging over mean motions. Mean motion resonances between the planets can introduce small divisors leading to divergences when forming the secular disturbing function. Secondly, there can be resonances between the proper mode frequencies leading to problems when one tries to solve the secular system using an expansion approach.

Laskar (1989) performed a critical test of the quasi-periodic hypothesis by numerically integrating the perturbations calculated to second order in mass and fifth order in eccentricities and inclinations. Such an expansion consists of about 150,000 polynomial terms. By numerically integrating the secular system, he avoids the small divisor problem caused by resonances between proper modes. The Fourier analysis of this 200 million year integration showed that it was not possible to describe the solution as a sum of periodic terms. Laskar also estimated the maximum Lyaponov exponent by the divergence of nearby orbits. The Lyaponov exponent should be zero for quasi-periodic orbits. Instead, he found the surprisingly high value of 1/5 Myr^{-1}. Figure 6 reproduces his plot of $\log(\gamma)$ vs $\log(t)$ in the notation of Section 2. In a subsequent paper, Laskar (1990) argued that the exponential divergence is due to the transition from libration to circulation of the critical argument of a secular resonance related to the motions of perihelia and nodes of the Earth and Mars. He argued from his results that the chaotic nature of the inner Solar System is robust against small variations in the initial conditions or in the model. These very important conclusions have recently been checked by direct numerical calculations (see below).

The analytical complexity of the perturbation techniques and the development of ever faster computers has led others to the investigation of stability by purely numerical models. The first numerical computation of planetary orbits was by Eckert et al (1951) who did a simulation of the outer planets for 350 years. This was extended by Cohen & Hubbard (1965) and Cohen et al (1973) to 120,000 years and 1 million years respectively. These integrations compared well with the perturbation calculations of Brouwer & van Woerkom, showing quasi-periodic behavior for the four major outer planets. Pluto's behavior, however, was sufficiently different to inspire further study. Williams & Benson (1971) performed a 4.5 million year integration of the secular motion of Pluto under the influence of the four Jovian planets. The motion of the Jovian planets were determined from Brouwer & van Woerkom's (1950) analytic solution. They found that the angle $3\lambda - 2\lambda_N - \bar{\omega}$ was in libration with a period of 20,000 yr, where λ and λ_N are the mean longitudes of Pluto and Neptune, respectively, and $\bar{\omega}$ is the longitude of perihelion of Pluto. As well, they found that the argument of perihelion of Pluto librates with a period of 4 Myr and that the angles $\Omega - \Omega_N$ and $\bar{\omega} - \bar{\omega}_N$ seemed to be in resonance with this libration. All these resonances acted to prevent close encounters of Pluto with Neptune and hence protect the orbit of Pluto.

The longest numerical integration done with conventional integration

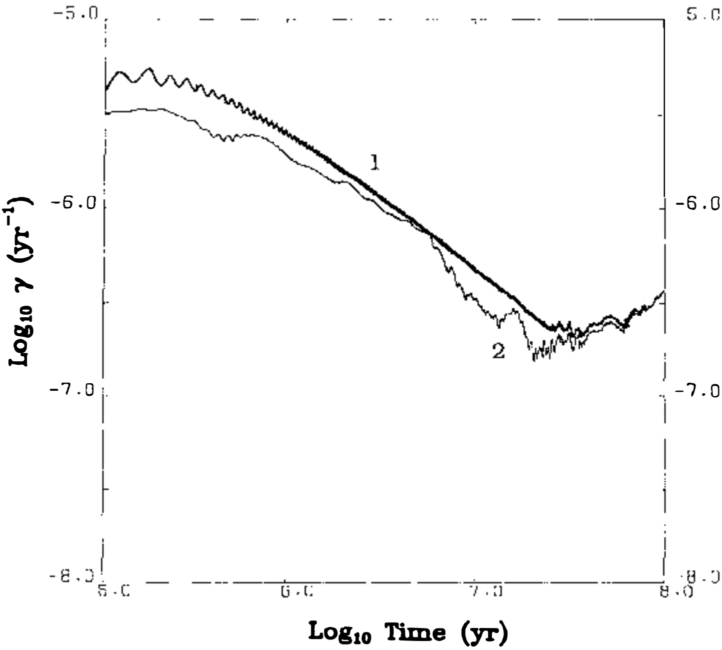

Figure 6 A plot of log(γ) versus log(t) (t in years) for the entire Solar System in the 100 Myr simulation of Laskar (1989). The two curves differ slightly because of differing renormalization procedures, but suggest the same Lyapunov timescale of 5 Myr.

schemes on general purpose computers was the LONGSTOP 1B integration done by Nobili et al (1989) which ran for 100 Myr. This integration followed the mutual interactions of the 5 outer planets, but also included the secular effects of the inner planets and the effects of general relativity. They have shown that these latter effects are significant over the time span of their integrations. Spectral analysis of their results suggested but did not prove that the orbits are chaotic. The evidence for this is regions of the spectrum where many lines of comparable amplitude accumulate, indicating that the series of Fourier terms is not converging.

Integrations of the outer planets for period up to 845 Myr or 20% of the age of the Solar System have been done with a special purpose machine: the Digital Orrery (Applegate et al 1986, Sussman & Wisdom 1988). This machine consists of one CPU per planet, with all the CPUs arranged in a ring. A force evaluation for all planets is accomplished in $\mathcal{O}(N)$ time instead of the usual $\mathcal{O}[N(N-1)/2]$, and the integration is performed in $\mathcal{O}(1)$ time. A disadvantage of using such a machine is that the nature of the force calculations is hard wired and the inclusion of effects such as general relativity could not be accomplished without rebuilding the machine. The dynamical system investigated by the Orrery is therefore slightly different than that investigated by Nobili et al. The longest of the integrations performed on the Orrery show that Pluto's orbit is not quasi-periodic. There is evidence for the existence of very long period changes in Pluto's orbital elements and Sussman & Wisdom (1988) calculate a Lyapunov exponent of $1/20\ \mathrm{Myr}^{-1}$. A subsequent reexamination of the LONGSTOP 100 Myr data led Milani et al (1989) to suggest that Pluto is locked in a complicated system of three resonances and that the value of the Lyapunov exponent for its motion could be sensitive to the assumed initial conditions and planetary masses. Thus, a detailed understanding of Pluto's behavior is likely to be obtained only with the next generation of simulations.

Until very recently, the length of direct integrations was limited, not by CPU time, but by roundoff error. Milani & Nobili (1988) concluded that it was impossible to reliably integrate the orbits of the outer planets for a period of 10^9 yr or more with the current computer hardware and software available at that time. The main problems they found were limited machine precision, and the empirical evidence that the longitude error after n steps was proportional to n^2 instead of $n^{3/2}$, as expected if the roundoff errors were uncorrelated. However, Quinn & Tremaine (1990) have proposed several corrections to the integration algorithms which considerably reduce the roundoff error, and Quinlan & Tremaine (1990) have proposed a high-order symmetric scheme. Taken together, these two techniques appear to maintain a longitude error growth that is linear in the number of steps and which may therefore allow for accurate integrations of the planets for tens of billions of years.

Quinn et al (1991) and Laskar et al (1992) used these newer integrators to make an accurate integration of of all nine planets and the Earth's spin axis for 3.05 Myr into the past, and future. Previously, Richardson & Walker (1987, 1989) performed 1 and 2 Myr integrations of all nine planets, but they neglected the rather important effects (for the inner planets) of General Relativity and the finite size of the Earth-Moon system. Quinn et al computed the long-term variations (periods >2000 yr) of the orbital elements of all the planets and the Earth's spin direction. These can be used to check or replace the results of secular perturbation theory and as input into geophysical models that test the Milankovich hypothesis that climate variations are caused by changes in the Earth's orbit. All of the planetary orbits appear to be regular over the 6 Myr integration span; however, this integration was not long enough to detect the chaotic motion found by Laskar. Comparison of the two simulations by Laskar et al (1992) show that they are in remarkable agreement over their common range. In particular, the numerical integrations demonstrated the same secular resonance which Laskar claims is responsible for the chaos seen in his longer simulation. The largest difference in these two solutions is in the eccentricity and inclination of Saturn which differed by at most 10%. This is as expected, since Saturn is near a 5:2 mean motion resonance with Jupiter. Such a resonance introduces small divisors into the higher order terms of the perturbation expansion and makes them large.

Another promising avenue for certain Solar System integrations are the symplectic schemes designed specifically to maintain the Hamiltonian structure of such systems of equations. General purpose symplectic integrators (Gladman et al 1991, Yoshida 1990) tend to be of low order because of their complexity and so are not suitable for long accurate simulations. However, so-called mixed variable symplectic (or MVS) integrators (Wisdom & Holman 1991, Saha & Tremaine, 1992) can be made more accurate by factors of the ratio of planetary to solar mass for a given timestep. The principle behind these integrators is to split the Hamiltonian into an unperturbed Kepler part and a perturbation part. In each step of the integration, the system is first moved forward in time according to Kepler motion, and then a kick in momentum is applied which is derived from the perturbation part of the Hamiltonian. This second step is analytic since the perturbed part of the Hamiltonian can be made independent of the canonical momenta in Cartesian coordinates. The MVS integrators have the additional advantage that the errors are limited to high frequency terms. Over long integration periods these terms will then average out, giving no net contribution to the evolution.

Sussmann & Wisdom (1992) used an MVS integrator on a somewhat special purpose computer to perform a 100 Myr integration using the same physical model and initial conditions as Quinn et al (1991) except for the treatment of General Relativity. Sussmann and Wisdom use the potential approximation of Nobili & Roxburgh (1986) which facilitates the use of the MVS integrator. Comparison of the two simulations gives a maximum difference in eccentricity of 1 part in 500 for Saturn. Again, this is expected from Saturn's mean motion resonance, since the mean motions depend on high frequency terms and are not calculated as accurately by the MVS integrator.

Sussmann and Wisdom's integration is actually 8 integrations, each with slightly different initial conditions, so that exponential divergence could be detected. In looking at the full secular phase space they found an initial exponential divergence with a timescale of 12 Myr, but after 60 Myr the divergence is dominated by an exponential with a timescale of 4 Myr. The 4 Myr divergence is in agreement with Laskar's estimate of 5 Myr, and, because of the slow convergence of Lyaponov estimates, the 12 Myr divergence is in agreement with the 15 to 20 Myr timescale divergence that they found in the Digital Orrery integrations. In examining divergences in individual planetary orbits, they found that the 4 Myr divergence occurs much later in the outer planets than in the inner planets. This seems to indicate that there are two distinct mechanisms generating divergence, and the inner planets are more sensitive indicators of the divergence with the shorter timescale. Sussmann & Wisdom are quite cautious about trying to identify the underlying dynamical mechanism causing this chaos. They note that the alternation between libration and circulation of a given angle is not necessarily indicative of a dynamically significant separatrix. Such behavior can also arise as an artifact of the projection of a high-dimensional trajectory onto a plane. The assignment of the responsibility for the chaotic behavior of the Solar System to a particular secular resonance will have to wait for an analytical demonstration that the resonances involved are sufficiently strong and close for resonance overlap.

Susmann & Wisdom also ran a 1 Gyr simulation of just the massive outer planets with their MVS integrator. They found the surprising result of a divergence with a 5 Myr timescale. Upon further checking with a traditional integrator, they confirmed the existence of this divergence, except that the timescale is closer to 20 Myr. Subsequent integrations with the MVS integrator with different step sizes showed divergence timescales varying from 3 Myr to 30 Myr. This spread in estimates appears to be due to the high frequency terms that are introduced in the MVS integrator, and which change according to the stepsize. Nevertheless, there is clear evidence that the outer massive planets are chaotic of themselves. Furthermore, they computed the divergence in the trajectories of massless Plutos in each of the outer planet integrations. This divergence always had a timescale between 10 and 20 Myr, irrespective of the timescale for divergence in the Jovian planets. The mechanism generating the chaotic behavior in Pluto appears to be independent of the chaos in the rest of the Solar System.

The detection of such large Lyapunov exponents indicates chaotic behavior. However, the apparent regularity of the motion of the Earth and Pluto and indeed the fact that the Solar System has survived for 4.5 billion years implies that the chaotic regions must be narrow. What the chaotic motion does mean (if confirmed) is that there is a horizon of predictability for the detailed motions of the planets. Thus, the exponential divergence of orbits with a 4–5 Myr timescale shown by Laskar and confirmed by Sussmann & Wisdom means that an error as small as 10^{-10} in the initial conditions will lead to a 100% discrepancy in 100 Myr. It is also worth bearing in mind the lessons learned from integration of test particle trajectories, namely that the timescale for macroscopic changes in the system can be many orders of magnitude longer than the Lyapunov timescales.

5.2 Other Planetary Systems

There are as yet no direct visual observations of planets around other stars, but there has been much recent interest in the millisecond pulsar PSR1257+12 because of the claim by Wolszcan & Frail (1992) that quasi-periodic variations in pulse arrival times are most straightforwardly interpreted as arising from perturbations from two planetary companions. Fits to the residuals yield orbital eccentricities for the planets of about 0.02, with orbital periods near to a 3:2 commensurability and masses of approximately $[3/\sin(i)]$ Earth masses, where i is the inclination of the orbit to that of the sky. This system proves to be very interesting dynamically (Rasio et al 1992, Malhotra et al 1992) and the orbital dynamics are reviewed by Malhotra (1992). In particular, dynamical stability arguments like those discussed in earlier sections can be used to place upper limits on the planetary masses, and the near-commensurability induces variations in the planetary orbital elements that should be detectable in the next few years.

One can of course construct initial conditions appropriate to *fictitious* planetary systems and numerically investigate the evolution and stability of the orbits. Early experiments of this type include those of Hills (1970), who argued on the basis of eleven systems integrated for a few thousand years that near-commensurability among the final orbital periods might represent a more stable end-state. It seems unlikely that the systems he investigated would remain stable over much longer timescales, however. Other early numerical and analytic work on the issue of the stability of the general three-body problem are reviewed in Message (1984).

More recent simulations along these lines have been performed by Quinlan (1992), who integrated the orbits on Myr timescales of systems with four Jovian planets. He found that the majority of fifty systems with initial conditions chosen at random from a reasonable probability distribution had Lyapunov timescales less than 10^5 years and that changes as small as 1% in the semimajor axis of any one of the planets from its real value could lead to chaotic motion on the timescale investigated. This superficially suggests that our own system is perhaps less chaotic than might have been expected and (if these results are extended by more experiments) requires an explanation.

6. SUMMARY

We have shown how the results from research in nonlinear Hamiltonian systems can be used to help us understand the recent results pertaining to the long-term dynamical evolution of orbits in the Solar System. Although algebraic mappings, analytic approaches, and perturbation theory have greatly aided our intuition in such nonlinear problems, much of our understanding usually comes initially from numerical simulations. We have seen that it is now possible to perform accurate integrations of planetary and test particle orbits on 10^8 year timescales throughout the Solar System and on 10^9 year timescales in the region of the giant planets. It is only a matter of time before these limits extend to the age of the Solar System. However, these investigations have become probabilistic in nature, so much remains to be done in the way of numerical exploration of the sorts of outstanding problems described in the sections above. Taken together, these investigations should lead to a better understanding of the origin, nature, and consequences of chaos in the long-term dynamics of the existing system.

It is also clear from what has been discussed here and elsewhere [see e.g. the review by Lissauer (1993) in this volume and also Wetherill (1991)] that long-range gravitational interactions must have influenced planet formation itself. It is now possible to undertake more realistic simulations of the late stages of planet formation, including the long-range mutual interactions of the planetary embryos. This work, together with further numerical experiments and analytic approaches of the sort described above, should deepen our understanding of the newly emerging picture of the dynamical evolution of our Solar System.

ACKNOWLEDGMENTS

The authors acknowledge helpful discussions with B. Gladman, H. Levison, J. Lissauer, R. Mulhotra, G. Quinlan, A. Stern, G. Stewart, and S. Tremaine . We also are grateful to those who commented on earlier drafts of this review, especially M. Bailey and P. Weissman. MJD gratefully acknowledges the hospitality of ITP (through NSF Grant PHY89-04035) where much of this review was written, and would like to thank Queen's University and the Natural Sciences and Engineering Research Council of Canada for continuing financial support.

Literature Cited

Applegate, J. H., Douglas, M. R., Gürsel, Y., Sussman, G. J., Wisdom, J. 1986. *Astron. J.* 92: 176–94
Arnold, J. R. 1965. *Ap. J.* 141: 1536–56
Arnold, V. I. 1961. In *Report to the IVth All-Union Mathematical Congress, Leningrad*, pp. 85–191
Bailey, M. E. 1983. *MNRAS* 204: 347–68
Bailey, M. E. 1986a. *Nature* 324: 350–52
Bailey, M. E. 1986b. *MNRAS* 218: 1–30
Bailey, M. E. 1991. In *Molecular Clouds*, ed. R. A. James, T. J. Millar, pp. 273–89. Cambridge: Cambridge Univ. Press
Bailey, M. E. 1992. *Celest. Mech.* In press
Bailey, M. E., Stagg, C. R. 1990. *Icarus* 86: 2–8
Benettin, G., Casartelli, M., Galgani, L., Giorgilli, A., Strelcyn, J.- M. 1978. *Nuovo Cimento* 44: 182–95
Berry, M. V. 1978. In *AIP Conf. Proc. No. 46, Topics in Nonlinear Dynamics*, ed. S. Jorna, E. C. Bullard, pp. 16–146. New York: Am. Inst. Phys.
Biermann, L. 1978. In *Astronomical Papers Dedicated to Bengt Strömgren*, ed. A. Reiz, T. Anderson, pp. 327–37. Copenhagen: Copenhagen Obs.

Bretagnon, P. 1974. *Astron. Astrophys.* 30: 141–54
Brouwer, D., Clemence, G. M. 1961. *Methods of Celestial Mechanics.* New York: Academic
Brouwer, D., van Woerkom, A. J. J. 1950. *Astron. Pap. Am. Ephem.* 13: part 2. Washington: US Print. Off.
Cohen, C. J., Hubbard, E. C. 1965. *Astron. J.* 70: 10–13
Cohen, C. J., Hubbard, E. C., Oesterwinter, C. 1973. *Astron. Pap. Am. Ephem.* 22: Part I. Washington: US Print. Off.
Delsemme, A. H. 1973. *Astron. Astrophys.* 29: 377–81
Dermott, S. F., Murray, C. D. 1983. *Nature* 301: 201–5
Duncan, M., Quinn, T. 1992. In *Protostars and Planets III*, ed. E. H. Levy, J. I. Lunine, M. S. Mathews. Tucson: Univ. Ariz. Press. In press
Duncan, M., Quinn, T., Tremaine, S. 1987. *Astron. J.* 94: 1330–38
Duncan, M., Quinn, T., Tremaine, S. 1988. *Ap. J. Lett.* 328: L69–73
Duncan, M., Quinn, T., Tremaine, S. 1989. *Icarus* 82: 402–18
Eckert, W. J., Brouwer, D., Clemence, G. 1951. *Astron. Pap. Am. Ephem.* 12. Washington: US Print. Off.
Everhart, E. 1972. *Ap. J. Lett.* 10: 131–35
Everhart, E. 1977. In *Comets—Asteroids—Meteorites*, ed. A. H. Delsemme, pp. 99–120. Toledo, Ohio: Univ. Toledo Press
Fernández, J. A. 1980. *MNRAS* 192: 481–91
Fernández, J. A., Ip, W.-H. 1983. *Icarus* 54: 377–87
Franklin, F., Lecar, M. 1992. *Astron. J.* In press
Franklin, F., Lecar, M., Soper, P. 1989. *Icarus* 79: 223–27
Froeschlé, C., Scholl, H. 1979. *Astron. Astrophys.* 72: 246–55
Gladman, B., Duncan, M. J. 1990. *Astron. J.* 100: 1669–75
Gladman, B., Duncan, M. J., Candy, J. 1991. *Celest. Mech.* 52: 221–40
Greenberg, R., Scholl, H., 1979. In *Asteroids*, ed. T. Gehrels, pp. 310–33. Tucson: Univ. Ariz. Press
Heisler, J. 1990. *Icarus* 88: 104–21
Heisler, J., Tremaine, S. 1986. *Icarus* 65: 13–26
Hénon, M. 1983. In *Chaotic Behaviour of Deterministic Systems*, ed. G. Iooss et al, pp. 54–169. Amsterdam: North-Holland
Hénon, M., Heiles, C. 1964. *Astron. J.* 69: 73–79
Heppenheimer, T. A. 1979. *Celest. Mech.* 20: 231–41
Hills, J. G. 1970. *Nature* 225: 840–42
Hills, J. G. 1981. *Astron. J.* 86: 1730–40
Hut, P., Tremaine, S. 1985. *Astron. J.* 90: 1548–57
Innanen, K. A. 1991. *J. R. Astron. Soc. Can.* 85: 151–57
Joss, P. C. 1973. *Astron. Astrophys.* 25: 271–73
Knezevic, Z., Milani, A., Farinella, P., Froeschlé, Ch., Froeschlé, Cl. 1991. *Icarus* 93: 316–30
Kowal, C. T. 1979. In *Asteroids*, ed. T. Gehrels, pp. 436–439. Tucson: Univ. Ariz. Press
Kuiper, G. P. 1951. In *Astrophysics: A Topical Symposium*, ed. J. A. Hynek, pp.357–424. New York: McGraw Hill
Laskar, J. 1989. *Nature* 338: 237–38
Laskar, J. 1990. *Icarus* 88: 266–91
Laskar, J., Quinn, T., Tremaine, S. 1992. *Icarus* 95: 148–52
Lecar, M., Franklin, F. 1973. *Icarus* 20: 422–36
Lecar, M., Franklin, F., Soper, P. 1992. *Icarus* 96: 234–50
Lemaitre, A., Henrard, J. 1990. *Icarus* 83: 391–409
Le Verrier, U.-J. 1856. *Ann. Obs. Paris* 2: 165–93
Levison, H. F. 1991. *Astron. J.* 102: 787–95
Levison, H. F., Duncan, M. J. 1990. *Astron. J.* 100: 1680–93
Levison, H. F., Duncan, M. J. 1993. *Ap. J. Lett.* 406: L35–38
Lissauer, J. J. 1993. *Annu. Rev. Astron. Astrophys.* 31: 129–74
Low F. et al 1984. *Astrophys. J. Lett* 278: L19–23
Luu, J., Jewitt, D. 1990. *Astron. J.* 100: 913–32
Luu, J., Jewitt, D. 1993. Preprint
Malhotra, R. 1992. In *Planets around Pulsars*, ed. J. A. Phillips, S. E. Thorsett, S. R. Kulkarni. In press
Malhotra, R., Black, D., Eck, A., Jackson, A. 1992. *Nature* 355: 583–85
Matese, J. J., Whitman, P. G. 1989. *Icarus* 82: 389–401
Meech, K., Belton, M. 1990. *Astron. J.* 100: 323–38
Message, P. J. 1984. *Celest. Mech.* 34: 155–63
Mikkola, S., Innanen, K. 1992. *Astron. J.* 104: 1641–49
Milani, A., Nobili, A. M., 1985. *Astron. Astrophys.* 144: 261–74
Milani, A., Nobili, A. M., 1988. *Celest. Mech.* 34: 343–55
Milani, A., Nobili, A. M., Carpino, M. 1989. *Icarus* 82: 200–17
Morris, D. E., Muller, R. A. 1986. *Icarus* 65: 1–12
Mumma, M. J., Weissman, P. R., Stern, S. A. 1992. In *Protostars and Planets III*, ed. E. H. Levy, J. I. Lunine, M. S. Mathews. Tucson: Univ. Ariz. Press. In press

Napier, W. M., Clube, S. V. M. 1979. *Nature* 282: 455–57
Nobili, A. M., 1989. In *Asteroids II*, ed. R. Binzel, T. Gehrels, M. S. Matthews, pp.862–79. Tucson: Univ. Ariz. Press
Nobili, A. M., Milani, A., Carpino, M. 1989. *Astron. Astrophys.* 210: 313–36
Nobili, A. M., Roxburgh, I. 1986. In *Relativity in Celestial Mechanics and Astrometry*, ed. A. Kovalevsky, C. Brumberg, pp. 20–29. Dordrecht: Reidel
Oikawa, S., Everhart, E. 1979. *Astron. J.* 84: 134–39
Oort, J. H. 1950. *Bull. Astron. Inst. Neth.* 11: 91–110
Öpik, E. J. 1951. *Proc. R. Irish Acad.* 54A: 165–99
Poincaré, H. 1892. *Les Methodes Nouvelles de la Méchanique Céleste*. Paris: Gauthiers-Villars
Quinlan, G. 1992. In *Chaos, Resonance and Collective Dynamical Phenomena in the Solar System*, ed. S. Ferraz-Mello, pp. 25–32. Amsterdam: Kluwer
Quinlan, G., Tremaine, S., 1990. *Astron. J.* 100: 1694–700
Quinn, T., Tremaine, S. 1990. *Astron. J.* 99: 1016–23
Quinn, T. R., Tremaine, S., Duncan, M. J. 1990. *Astrophys. J.* 355: 667–79
Quinn, T. R., Tremaine, S., Duncan, M. J. 1991. *Astron. J.* 101: 2287–305
Rasio, F. A., Nicholson, P. D., Shapiro, S. L., Teukolsky, S. A. 1992. *Nature* 355: 325–26
Richardson, D. L., Walker, C. F. 1987. In *Astrodynamics* 1987, ed. J. K. Soldner, A. K. Misra, R. E. Lindberg, W. Williamson, *Adv. Astron. Sci.* 65: 1473–95. San Diego: Univelt
Richardson, D. L., Walker, C. F. 1989. *J. Astronaut. Sci.* 37: 159–82
Saha, P., Tremaine, S., 1992. *Astron. J.* 104: 1633–40
Scholl, H., Froeschlé, Ch. 1974. Astron. Astrophys. 33: 455–58
Scholl, H., Froeschlé, Ch., Kinoshita, H., Yoshikawa, M., Williams, J. G. 1989. In *Asteroids II*, ed. R. P. Binzel, T. Gehrels, M. S. Matthews, pp. 845–61. Tucson: Univ. Ariz. Press
Schubart, J. 1968. *Astron. J.* 73: 99–103
Shoemaker, E. M., Wolfe, R. F. 1984. *Lunar Planet. Sci.* 25: 780–81 (Abstr.)
Smoluchowski, R., Torbett, M. 1984. *Nature* 311: 38–39
Stagg, C. R., Bailey, M. E. 1989. *MNRAS* 241: 507–41
Stern, A. 1991. *Icarus* 90: 271–81
Sussman, G. J., Wisdom, J. 1988. *Science* 241: 433–37
Sussman, G. J., Wisdom, J. 1992. *Science* 257: 56–59
Torbett, M. 1986. In *The Galaxy and the Solar System*, ed. M. S. Matthews, J. Bahcall, R. Smoluchowski, pp. 147–72. Tucson : Univ. Ariz. Press
Torbett, M. 1989. *Astron. J.* 98: 1477–81
Torbett, M., Smoluchowski, R. 1990. *Nature* 345: 49–51
Weinberg, M. D., Shapiro, S. L., Wasserman, I. 1986. *Icarus* 65: 27–36
Weinberg, M. D., Shapiro, S. L., Wasserman, I. 1987. *Ap. J.* 312: 367–89
Weissman, P. R. 1986. In *The Galaxy and the Solar System*, ed. M. S. Matthews, J. Bahcall, R. Smoluchowski, pp. 204–37. Tucson: Univ. Ariz. Press
Weissman, P. R. 1990. *Nature* 344: 825–30
Wetherill, G. W. 1990. In *Comets in the Post-Halley Era: IAU Colloq. No. 116*, ed. R. L. Newburn, J. Rahe, M. Neugebauer, pp. 537–56. Amsterdam: Kluwer
Wetherill, G. W. 1991. *Science* 253: 535–38
Whipple, F. L. 1964. *Proc. Natl. Acad. Sci. U.S.* 51: 711–18
Williams, J. G. 1969. *Secular Perturbations in the Solar System*. PhD dissertation. Univ. Calif., Los Angeles
Williams, J. G. 1971. In *Physical Studies of Minor Planets*, ed. T. Gehrels, pp. 177–181. NASA publ. SP-267
Williams, J. G., Benson, G. S. 1971. *Astron. J.* 76: 167–77
Wisdom, J. 1980. *Astron. J.* 85: 1122–33
Wisdom, J. 1982. *Astron. J.* 87: 577–93
Wisdom, J. 1983. *Icarus* 56: 51–74
Wisdom, J. 1985. *Icarus* 63: 272–79
Wisdom, J. 1987. *Icarus* 72: 241–75
Wisdom, J., Holman, M. 1991. *Astron. J.* 102: 1528–38
Wolszczan, A., Frail, D. A. 1992. *Nature* 355: 145–47
Yabushita, S., Tsujii, T. 1991. *MNRAS* 252: 151–55
Yoshida, H. 1990. *Phys. Lett. A* 150: 262–68
Yoshikawa, M. 1991. *Icarus* 92: 94–117