Symplectic Maps for the N-body Problem with Applications to Solar System Dynamics

by

Matthew Jon Holman
S.B. (Mathematics)
Massachusetts Institute of Technology (1989)

Submitted in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY
IN
PLANETARY SCIENCE
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1994

Author .....
Department of Earth, Atmospheric, and Planetary Sciences
April 22, 1994

Certified by .....
Jack Wisdom
Professor
Thesis Supervisor

Accepted by .....
Thomas Jordan
Department Head

Symplectic Maps for the N-body Problem with Applications to Solar System Dynamics

by

Matthew Jon Holman

S.B. (Mathematics)

Massachusetts Institute of Technology (1989)

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences
on April 22, 1994, in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in
Planetary Science

Abstract

The mapping method of Wisdom (1982) is generalized to encompass all gravitational n-body problems with a dominant central mass. The method is used to compute the evolution of the outer planets for a billion years, providing independent numerical confirmation of the result of Sussman and Wisdom (1988) that the motion of the planet Pluto is chaotic.

The stability of the symplectic mapping method for the n-body problem introduced recently by Wisdom and Holman (1991) is analysed in a novel application of the methods of non-linear dynamics.

Test particle stability in the outer solar system is surveyed. Clusters of test particles near the triangular Lagrange points of Jupiter, Saturn, Uranus, and Neptune survive the full integration, here 20 million years. Nearly all particles started on circular orbits between the outer planets are removed by close encounters with the planets during the course of 4.5 billion years integrations. Numerous test particles between Neptune and 43 AU are removed by close encounters with Neptune. The distribution of encounter times suggests that the times to first encounter can reach several billion years. The flux of new encounters decays slowly, roughly as the inverse of time. An estimate of the mass of the Kuiper belt is given.

Thesis Supervisor: Jack Wisdom
Title: Professor

Acknowledgments

I am grateful to Jack Wisdom for his support and guidance throughout my graduate studies. When I began graduate school under Jack's supervision I was unaware of the enormous opportunity that had been given to me. It is a pleasure to thank Gerald Sussman for many helpful, friendly conversations (conversions) that have sparked creative ideas. I am grateful to Jihad Touma for the kind of intellectual and spiritual support that only fellow graduate students can provide. Also, I thank Richard Binzel for his efforts to keep my graduate career moving forward. I would like to acknowledge Mireille Dailey for her assistance in conducting the test particle surveys. Kimmo Innanen, Gerald Quinlan, and Scott Tremaine made several helpful suggestions. I thank my family, Amanda Bosh, and many friends for their support over the last few years. Last, I thank James Elliot for his laboratory course in optical astronomy which initially lured me into planetary science.

This work was supported in part by an NSF Graduate Fellowship, by an NSF Presidential Young Investigator Award (AST-8857365), with an industrial match from the Hewlett-Packard Corporation, and by the NASA Planetary Geology and Geophysics Program (NAGW-706).

Contents

[1] Adapted from Wisdom, J. and M. Holman (1991). Symplectic Maps for the N-body Problem. *Astron. J.* **102**, 1528–1538.

[2] Adapted from Wisdom, J. and M. Holman (1992). Symplectic Maps for the N-body Problem: Stability Analysis. *Astron. J.* **104**, 2022–2029.

[3]Adapted from Holman, M. and J. Wisdom (1993). Dynamical Stability in the Outer Solar System and the Deliver of Short Period Comets. *Astron. J.* **105**, 1987–1999.

List of Figures

List of Tables

Chapter 1

Introduction

Long-term numerical integration is a mainstay of solar system dynamics. Many important questions about dynamical phenomena in the solar system cannot be answered analytically. In fact, careful numerical exploration has revealed many surprising results. For example, the 100 million year integrations of the full solar system carried out by Sussman and Wisdom (1993) provide strong evidence that the motion of the system of planets is chaotic. That is to say, model solar systems with small differences in initial conditions divergence exponentially from each other on a timescale of 10–20 million years. In another example, Touma and Wisdom (1993) show with 10 million year integrations of the rigid-body dynamics of Mars in the field of the Sun and other planets that the obliquity of Mars also varies chaotically.

Typically, numerical integrations of solar system problems that are long enough to reveal the relevant dynamical characteristics are computationally expensive. Therefore, the development of efficient and accurate algorithms for integrating the solar system n-body problems is important. The symplectic mapping for the n-body problem, described in this work, is such an algorithm. This technique is designed to exploit two aspects of the solar system n-body problem. First, the method relies on the fact that the orbits of the planets about the Sun are basically Keplerian ellipses. Second, the method is based upon the Hamiltonian formulation of the n-body problem and, unlike conventional integration techniques, preserves the Hamiltonian characteristics of the system. These two points contribute to a method that is typically an order

of magnitude faster than conventional methods. Much of the research completed using the symplectic mapping method would be impossible to complete without this speed. For example, the 100 million year integration of the solar system mentioned above required approximately one thousand hours or about one month on a computer dedicated to that problem. Without the symplectic mapping method this same computation would have taken ten months.

Aside from speed, another advantage of the symplectic mapping method for the n-body problem is that the algorithm itself can be described as a Hamiltonian dynamical system. Therefore, it is possible to analyze the stability of the method using standard techniques of modern non-linear dynamics. This stability analysis is developed herein.

The speed and stability properties of the symplectic mapping method for the n-body problem permit studies of solar system dynamics with broad scope. One such application is a survey of test particle stability in the outer solar system. The goal here is to examine the long-term stability of small bodies started on circular orbits in the plane of the solar system against close encounters with the planets. With this type of study we can determine if regions exist between the giant planets where small bodies can persist with low-eccentricity orbits for the age of the solar system. Furthermore, we can determine which regions the bodies are removed from and the timescales on which these removal processes occur.

The remainder of the work is organized as follows: Chapter 2 develops the symplectic mapping method for the n-body problem. Chapter 3 presents a stability analysis of the mapping method based on the resonance overlap criterion. Chapter 4 describes the application of the symplectic mapping method to a survey of test particle stability in the outer solar system. Chapter 5 summarizes the contents of this thesis and describes directions for future work.

Chapter 2

Symplectic Maps for the N-Body Problem[1]

2.1 Introduction

Long-term integrations are playing an increasingly important role in investigations in dynamical astronomy. The reason is twofold. First, numerical exploration is an essential tool in the study of complex dynamical systems which can exhibit chaotic behavior, and there has been a growing realization of the importance of chaotic behavior in dynamical astronomy (see e.g. Wisdom, 1987). Second, there has been a phenomenal increase in the capabilities of computers which is bringing many important problems in dynamical astronomy within reach. In particular, there has recently been considerable interest in the long-term evolution of the solar system. Long-term integrations of the solar system include the outer planet integrations of Cohen, Hubbard, and Oesterwinter (1973; 1Myr), Kinoshita and Nakai (1984; 5Myr), the first Digital Orrery integration (Applegate et al. 1986, 210Myr), the LONGSTOP work (Roy et al. 1988; 100Myr), the second Digital Orrery integration (Sussman and Wisdom, 1988, 845Myr), and the inner planet integrations of Richardson and Walker (1987; 2Myr), Applegate et al. (1986, 3Myr), and Quinn, Tremaine, and Duncan (1991; 3Myr). Long-term integrations have already produced startling results. Sussman and Wisdom (1988) found numerical evidence that the motion of the planet

[1]Adapted from Wisdom, J. and M. Holman (1991). Symplectic Maps for the N-body Problem. *Astron. J.* **102**, 1528–1538.

Pluto is chaotic, with a remarkably short timescale for exponential divergence of trajectories of only 20 million years. This massive calculation consumed several months of time on the Digital Orrery, a computer built specifically for the job which runs at about a third the speed of a Cray 1. Subsequently, Laskar (1989, 1990), in another massive computation, found numerical evidence that the motion of the inner planets is also chaotic, with a divergence timescale of only 5 million years. However, despite the phenomenal progress in computer technology, computers are still too slow for many important applications. For example, it is very important to test the sensitivity of the results concerning the chaotic character of the motions of the planets to uncertainties in initial conditions and parameters. It is also important to clarify the dynamical mechanisms responsible for the chaotic behavior to confirm that the positive Lyapunov exponents are not subtle numerical artifacts. The necessary calculations and those of many other problems of current interest in dynamical astronomy require orders of magnitude greater computing power than is currently available. Regardless of the speed of computers, better, faster algorithms for investigating the n-body problem are always welcome. This paper presents a new method for studying the long-term evolution of the n-body problem which is an order of magnitude faster than traditional methods of numerical integration. The method is a generalization of the "mapping" method introduced by Wisdom (1982, 1983) to study the motion of asteroids near the 3:1 mean-motion resonance with Jupiter. It is applicable to systems which are dominated by a large central mass such as planetary systems or satellite systems.

The mapping method of Wisdom (1982, 1983) was based on the averaging principle. It was noted that most studies of the long term evolution of the n-body problem relied on the averaging principle in one way or another. This included both analytical and numerical studies. The intuition behind the averaging method is that rapidly oscillating terms tend to average out and give no net contribution to the evolution, while more slowly varying resonant or secular terms accumulate to give significant contributions to the evolution (see Arnold, 1978). The intuition behind the mapping method was just the same: If the rapidly oscillating terms do not contribute signifi-

cantly to the evolution then replacing them with other rapidly oscillating terms will have no ill effect. To get the mapping the rapidly oscillating terms are chosen so that they sum to give delta functions which can be locally integrated to give explicit equations specifying how the system changes from one step to the next. The mapping method was inspired by Chirikov's use of periodic delta functions to derive a Hamiltonian for the standard map (Chirikov, 1979). The time step covered by the map is on the order of the period associated with the high frequency terms. For the asteroid maps the basic step was one full Jupiter period. The algebraic simplicity of the 3:1 map and the large step-size combined to make it extraordinarily fast, about 1000 times faster than even the numerical averaging routines available at the time (Wisdom, 1982). The great speed of the map allowed studies of the resonant asteroid motion over much longer times than were previously possible, and significant new phenomena were discovered. In particular, it was found that there was a large zone of chaotic behavior near the 3:1 resonance and that chaotic trajectories in these zones often displayed a peculiar phenomena in which the eccentricity could remain at relatively low values for several hundred thousand years and then suddenly jump to much higher values. Over longer intervals of millions of years there were periods of low eccentricity behavior interspersed with bursts of high eccentricity behavior. These bursts in eccentricity were subsequently confirmed in traditional direct integrations of Newton's equations (Wisdom, 1983, Murray and Fox, 1984, Wisdom, 1987), and explained perturbatively (Wisdom, 1985a). The high eccentricities attained by the chaotic trajectories help explain the formation of the 3:1 Kirkwood gap (Wisdom, 1983), as well as provide a mechanism for transporting meteoritic material directly from the asteroid belt to Earth (Wisdom, 1985b, Wetherill, 1985). Murray (1986) applied the mapping method to the 2:1 and the 3:2 resonances. Sidlichovsky and Melendo (1986) applied the method to the 5:2 resonance. Tittemore and Wisdom (1988, 1989, 1990) have applied the method to study the tidal evolution of the Uranian satellites through numerous mean-motion commensurabilities. The result of Tittemore and Wisdom (1989) that secondary resonances play a crucial role in determining the inclination of Miranda has been confirmed by Malhotra and Dermott (1990), also

using the mapping method. Tittemore (1990) and Malhotra (1990) have recently used the mapping method to study the tidal evolution of the Gallilean satellites. The mapping method has been tremendously useful.

Unfortunately, the mapping method, as originally presented, has significant limitations. It is based on analytic representations of the averaged Hamiltonian near particular resonances. The only known explicit analytic representations of the averaged disturbing function are as expansions in the eccentricities and inclinations, or the canonical equivalents. Though the mapping method itself has no particular limitation to low eccentricities and inclinations, the use of a disturbing function which is truncated at some order in both eccentricity and inclination limits the applicability of any particular realization of the mapping method to low eccentricity and inclination. The original 3:1 mapping which included second order terms in eccentricity and inclination (ignoring fourth order terms) gave qualitatively correct trajectories even for eccentricities as large as 0.4. However, it could not be relied upon for the investigation of Earth-crossing meteoroid trajectories which have eccentricities above 0.6 (Wisdom, 1985). Murray (1986) also used a disturbing function truncated after second order terms in the eccentricity in his study of motion near the 2:1 and the 3:2 resonances (ignoring third order terms). The eccentricity must be much smaller at the 2:1 and the 3:2 resonances than at the 3:1 resonance for a second order disturbing function to accurately represent the motion. Comparison of his results with those obtained with unaveraged numerical integrations performed on the Digital Orrery (Wisdom, 1987) show that significant artifacts appear in Murray's maps above an eccentricity of only 0.1 at the 2:1 resonance, and the extent of the chaotic regions determined by Murray's map is qualitatively wrong for the 3:2 resonance even at low eccentricity. It is important to emphasize that the failure of Murray's maps is not a failure of the mapping method, but rather a failure of the truncated disturbing function to represent the averaged Hamiltonian. Another important limitation of the mapping method, as it has been used up to the present, is that it is limited to the vicinity of a particular resonance or group of resonances, again because the analytic representation of the averaged disturbing function can only be made for a particular

set of commensurabilities. Thus the systematic investigation of the tidal evolution of the Uranian satellite system through a sequence of mean-motion commensurabilities (Tittemore and Wisdom, 1988, 1989, 1990) required a separate derivation of the map appropriate to each resonance, an unbelievably tedious process!

The generalization of the mapping method presented here does not have these limitations. It is not limited to particular resonances, nor is it limited to low eccentricities and inclinations. It is valid everywhere. Of course this comes at a cost. The new mapping method is not as fast as the original mapping method, but it still offers a significant advantage over conventional direct numerical integration.

The next section presents the rationale for the generalized mapping method. Details of the mapping for the n-body problem are then presented. Subsequent sections present some refinements of the method and show the relationship of the mapping method to other symplectic integration methods. The new n-body map has been used to compute the evolution of the outer planets for a billion years. The resulting evolution is compared to the 845 million year evolution of the outer planets performed on the Digital Orrery using standard numerical integration techniques (Sussman and Wisdom, 1988).

2.2 Mapping Method

In the original mapping method the Hamiltonian is first separated analytically into parts with different associated timescales,

$$H = H_{Kepler} + H_{Orbital} + H_{Resonant} + H_{Secular}, \tag{2.1}$$

where H_{Kepler} represents the interaction of each body with the central mass, $H_{Orbital}$ represents rapidly oscillating terms which depend on the mean longitudes of the bodies but are not resonant in the region of interest, $H_{Resonant}$ represents the terms which have resonant combinations of mean longitudes, and $H_{Secular}$ represents the remaining terms which do not depend on mean longitudes. The averaging principle is used to argue that the terms in $H_{Orbital}$ will not significantly affect the long-term evolution

near resonance, and can thus be neglected (or removed by suitable Von Zeipel transformations). The original mapping method then added additional terms with the orbital frequency, which sum, together with $H_{Resonant}$, into terms involving periodic sequences of Dirac delta functions. The new terms, by the averaging principle, also play no important role in the long-term evolution. The resulting map Hamiltonian is

$$H_{Map} = H_{Kepler} + H_{Secular} + 2\pi\delta_{2\pi}(\Omega t)H_{Resonant}, \tag{2.2}$$

where $\delta_{2\pi}(t)$ represents a periodic sequence of delta functions with period 2π,

$$\delta_{2\pi}(t) = \sum_{n=-\infty}^{\infty} \delta(t - 2\pi n) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} \cos(nt), \tag{2.3}$$

and Ω is the mapping frequency, which is of the same order as the orbital frequencies. In the asteroid map, the mapping period was chosen to be the period of Jupiter. Hamiltonian (2.2) is only a sketch of the true mapping Hamiltonian because in the earlier applications it was convenient to break $H_{Resonant}$ into several parts, each of which was multiplied by its own sequence of delta functions. Those details are not important here. Between the times when the delta functions act, the Hamiltonian is just given by the first two parts, the Kepler part and the secular part. Provided the secular Hamiltonian is truncated at second order in eccentricities and inclinations (ignoring fourth order terms), Hamilton's equations can be solved analytically between the delta functions. The system can also be analytically integrated across the delta functions. The result is an analytic expression for the state of the system at the end of a mapping period in terms of the state of the system at the beginning of the mapping period. The time evolution of the system is obtained by iterating the mapping step. It is easily shown that the mapping step is a canonical transformation or, in other words, that the mapping is symplectic. Again, these mappings are limited to particular regions in which only certain resonant terms are important, and to low eccentricity and inclination by the truncation of both $H_{Secular}$ and $H_{Resonant}$ to some manageable order.

The mappings presented in this paper are based on a simpler separation of the Hamiltonian for the n-body problem:

$$H = H_{Kepler} + H_{Interaction}, \tag{2.4}$$

where again H_{Kepler} represents the basic Keplerian motion of the bodies with respect to the central body, and $H_{Interaction}$ represents the perturbation of the bodies on one another. Of course, this division of the Hamiltonian is very natural and is the starting point for most perturbation theory. Despite this, few numerical integration methods take advantage of this division of the problem. Encke's method makes use of the integrability of the Kepler problem by integrating the variations of the planetary trajectories with respect to fixed reference orbits. However, a serious problem with Encke's method is that as the system evolves new reference trajectories must be frequently chosen so that the variations are not too large, and the solutions for different reference trajectories must be smoothly connected. Nevertheless, Encke's method was successfully used in the LONGSTOP integrations (Roy et al. 1988) to reduce numerical error. The symplectic n-body maps introduced here are quite distinct from Encke's method and more fully exploit the integrability of the Kepler problem.

A mapping Hamiltonian for the n-body problem can be simply obtained by adding high frequency terms to this Hamiltonian so that it becomes

$$H_{Map} = H_{Kepler} + 2\pi\delta_{2\pi}(\Omega t)H_{Interaction}. \tag{2.5}$$

More refined versions of the mapping Hamiltonian will be presented in subsequent sections. In all of these mapping Hamiltonians high frequency terms are introduced without first removing terms of corresponding frequency from the Hamiltonian. Nevertheless, by the averaging principle, the new high frequency terms are unimportant. Our new n-body maps then consist of a sequence of steps alternating pure Keplerian evolution of the individual bodies between the delta functions, with periodic interaction kicks derived from integrating the whole system across the delta functions. The basic idea is remarkably simple.

The construction of an efficient mapping for any problem rests on the ability to separate the Hamiltonian into parts which are themselves not only integrable, but efficiently computable. This looks grim at first sight for this problem. Keplerian motion is integrable, but the solution is naturally expressed only in terms of Keplerian orbital elements or one of the canonical equivalents such as the Delaunay variables.

The description of the gravitational interaction of two bodies in terms of Keplerian orbital elements leads again to the expansion of the disturbing function with all the attendant complications and limitations. A map of this form would be useless. However, there is no particular reason to insist on one single set of coordinates. In fact, since the evolution for each part of the Hamiltonian is computed separately, each can be evaluated in the coordinates most suitable for that part: the Kepler orbits can be advanced in canonical Keplerian elements, and the interactions can be evaluated in canonical Cartesian coordinates, with of course the appropriate intermediate canonical transformations. There is still a better solution. The Kepler orbits can be advanced directly in canonical Cartesian coordinates using Gauss' f and g functions (see Danby, 1988) without ever having to convert to Keplerian elements. This can be naturally combined with kicks resulting from the interaction Hamiltonian evaluated directly in canonical Cartesian coordinates. It is amusing that Cartesian coordinates appear to be the best coordinates to use to take full advantage of the fact that the basic motion is Keplerian.

2.3 N-Body Problem

The Hamiltonian for the n-body problem is

$$H = \sum_{i=0}^{n-1} \frac{p_i^2}{2m_i} - \sum_{i<j} \frac{Gm_i m_j}{r_{ij}}. \tag{2.6}$$

In order to make the n-body maps this must be separated into a Keplerian Hamiltonian and an interaction Hamiltonian. A Hamiltonian is Keplerian if it can be written in the form

$$H = \frac{p^2}{2m} - \frac{\mu}{r}, \tag{2.7}$$

or as a sum of such forms. Unfortunately, the n-body Hamiltonian is not immediately in the desired form. For the two-body problem, the separation of the Hamiltonian into a Kepler Hamiltonian and a non-interacting center of mass Hamiltonian is achieved by transforming to relative coordinates and center of mass coordinates. For the n-body problem, it is easy to show that a similar transformation to coordinates relative

to the central mass plus center of mass coordinates does not produce a Hamiltonian which is a sum of non-interacting planetary Kepler Hamiltonians, center of mass Hamiltonian, and an interaction Hamiltonian. The problem is that with this simple choice of relative coordinates the kinetic energy is no longer a diagonal sum of squares of the new momenta. Of course, the choice of variables which accomplishes the desired transformation of the n-body Hamiltonian is well known, and is just the Jacobi coordinates (see Plummer, 1960). The Jacobi coordinates can be derived by writing them as a general linear contact transformation, then requiring that the kinetic energy remain a diagonal sum of squares of the new momenta, and also that the new Hamiltonian be cyclic in the center of mass coordinate. The latter condition means that all the distances between the bodies can be written in term of $n-1$ of the new "relative" coordinates. Despite this formal motivation, the resulting Jacobi coordinates turn out to have a simple interpretation. We take the first coordinate to be the position of the center of mass. The first relative coordinate is just the position of the first planet relative to the central mass. The second relative coordinate is the position of the second planet relative to the center of mass of the central mass and the first planet. In general, the i^{th} relative coordinate is the position of the i^{th} planet relative to the center of mass of the central mass and the planets with lower indices. It is not necessary, but increasing indices are usually taken to correspond to increasing semi-major axes.

Denoting the Jacobi coordinates by a prime, the first Jacobi coordinate $\vec{x}_0{}'$ is the center of mass. The remaining $n-1$ Jacobi coordinates are $(0 < i < n)$

$$\vec{x}_i{}' = \vec{x}_i - \vec{X}_{i-1}, \tag{2.8}$$

where $\vec{X}_i$ denotes the center of mass of bodies with indices up to i.

$$\vec{X}_i = \frac{1}{\eta_i} \sum_{j=0}^{i} m_j \vec{x}_j, \tag{2.9}$$

with the definition

$$\eta_i = \sum_{j=0}^{i} m_i. \tag{2.10}$$

In terms of the $\vec{X}_i$, the first Jacobi coordinate is simply $\vec{x}_0{}' = \vec{X}_{n-1}$, the center of mass of the whole system. By virtue of the requirement that the new Hamiltonian is a diagonal sum of the squares of the new momenta, the momenta conjugate to the $\vec{x}_i{}'$ have the familiar form $\vec{p}_i{}' = m_i{}'\vec{v}_i{}'$, where $\vec{v}_i{}'$ is the time derivative of $\vec{x}_i{}'$. The new mass factors are given by $m_i{}' = \eta_{i-1}m_i/\eta_i$, for $0 < i < n$ and $m_0{}' = \eta_{n-1} = M$, the total mass of the system. It is only a matter of algebra to show that in terms of these Jacobi coordinates the Hamiltonian for the n-body problem becomes

$$H = \frac{p_0{}'^2}{2M} + \sum_{i=1}^{n-1} \frac{p_i{}'^2}{2m_i{}'} - \sum_{i=1}^{n-1} \frac{Gm_im_0}{r_{i0}} - \sum_{0<i<j} \frac{Gm_im_j}{r_{ij}}, \tag{2.11}$$

where $r_{ij} = ||\vec{x}_i - \vec{x}_j||$, the distance between bodies i and j. By construction, r_{ij} does not depend on $\vec{x}_0{}'$, thus the total momentum $\vec{p}_0{}'$ is an integral of the motion. As expected, the center of mass moves as a free particle. Hereinafter the center of mass contribution to the Hamiltonian will be omitted. Adding and subtracting the quantity

$$\sum_{i=1}^{n-1} \frac{Gm_im_0}{r_i{}'}, \tag{2.12}$$

where $r_i{}' = ||\vec{x}_i{}'||$ the Hamiltonian becomes

$$H = \sum_{i=1}^{n-1} \left(\frac{p_i{}'^2}{2m_i{}'} - \frac{Gm_im_0}{r_i{}'} \right) + \sum_{i=1}^{n-1} \left(\frac{Gm_im_0}{r_i{}'} - \frac{Gm_im_0}{r_{i0}} \right) - \sum_{0<i<j} \frac{Gm_im_j}{r_{ij}}, \tag{2.13}$$

The second sum, which we may call the indirect perturbation, contains differences of nearly equal quantities and is actually of the same order as the direct interaction terms. The Hamiltonian now separates into a sum of $n-1$ non-interacting Kepler Hamiltonians, and a smaller interaction Hamiltonian, as desired:

$$H = H_{Kepler} + H_{Interaction}, \tag{2.14}$$

with

$$H_{Kepler} = \sum_{i=1}^{n-1} \left(\frac{p_i{}'^2}{2m_i{}'} - \frac{Gm_im_0}{r_i{}'} \right), \tag{2.15}$$

and

$$H_{Interaction} = \sum_{i=1}^{n-1} Gm_im_0 \left(\frac{1}{r_i{}'} - \frac{1}{r_{i0}} \right) - \sum_{0<i<j} \frac{Gm_im_j}{r_{ij}}. \tag{2.16}$$

It is common to expand the interaction Hamiltonian in terms of the small differences between the Jacobi coordinates and the heliocentric Cartesian coordinates, and keep only the first order corrections in the ratio of the planetary masses to the mass of the central body. In this approximation the interaction Hamiltonian becomes

$$H_{Interaction} = -\sum_{0<i<j} \left[\frac{Gm_i m_j}{r_{ij}'} - \frac{Gm_i m_j \vec{x}_i' \cdot \vec{x}_j'}{r_j'^3} \right] + o(m_i^3), \tag{2.17}$$

with $r_{ij}' = ||x_i' - x_j'||$. We have found though that maps based on the exact Hamiltonian are nearly as efficient as those based on this approximate Hamiltonian, so the expanded form will not be considered further.

An important special case of the n-body problem is obtained if some of the bodies are given infinitesimal mass. These "test particles" are perturbed by the massive planets, but do not perturb them in return. The restricted three-body problem and all its variations such as the planar elliptic restricted three-body problem fall in this category. If the test particle is given the first relative Jacobi index, below those of the massive particles, then the test particle interaction Hamiltonian is given exactly by

$$H_{TestParticle} = -\sum_{j>1} \left[\frac{Gm_j}{r_{1j}} - \frac{Gm_j \vec{x}_1' \cdot \vec{x}_{j0}}{r_{j0}^3} \right], \tag{2.18}$$

where $\vec{x}_{j0} = \vec{x}_j - \vec{x}_0$, the vector from the central mass to body j. There are several ways of deriving this. The most straighforward method is to expand the exact interaction Hamiltonian in the differences between the Jacobi coordinates and heliocentric Cartesian coordinates, then take the appropriate limit as the test particle mass goes to zero. There are several alternate routes. The most intuitive is to note that the acceleration of the vector from the central mass to the test particle is the difference of the direct acceleration of the test particle and the acceleration of the central mass due to the gravitational attraction of the other massive bodies. This immediately gives the same test particle interaction Hamiltonian. If the test particle is given any other Jacobi index the interaction Hamiltonian is more complicated, and will not be given here. Though the equations of motion are simpler if the test particle is given a Jacobi index below the massive bodies, the orbital elements are "cleaner" if the test particle is given the natural Jacobi index in order of increasing semimajor axis along

with the massive planets (see Sussman and Wisdom, 1988). The resulting orbital elements are then freed of the relatively rapid oscillations due to the motion of the central mass induced by those massive planets interior to the test particle.

2.4 N-Body Maps

A simple mapping Hamiltonian for the n-body problem is then just Hamiltonian (2.5). It involves two distinct operations: advancing the Kepler orbits between the delta functions, and integrating the system across the delta functions. The more refined n-body maps to be presented later use the same components.

Constructing an efficient map depends on being able to rapidly advance Keplerian orbits. A summary of methods for solving this classic initial value problem is given in Danby (1988). A key element in the solutions is that the motion can be determined without explicitly determining the orientation of the orbit plane. In particular, since two vectors determine a plane, the position and velocity at any time can be written as a time dependent linear combination of the position and velocity at the initial epoch

$$\vec{x}(t) = f(t)\vec{x}(t_0) + g(t)\vec{v}(t_0) \tag{2.19}$$

and consequently

$$\vec{v}(t) = \dot{f}(t)\vec{x}(t_0) + \dot{g}(t)\vec{v}(t_0), \tag{2.20}$$

using Gauss' famous f and g functions. Refer to Danby for a derivation of the equations which determine f and g, and their time derivatives. An important step in the determination of f and g is the calculation of the change in the eccentric anomaly, ΔE. In this paper we concentrate on problems for which the time step is a fraction of an orbit period. For this case, it is efficient to determine ΔE through the solution of the difference form of Kepler's equation

$$\Delta M = n\Delta t = \Delta E - e\cos E_0 \sin\Delta E + e\sin E_0(1-\cos\Delta E), \tag{2.21}$$

where e is the orbital eccentricity, M is the mean anomaly, n is the mean motion, and E_0 is the initial eccentric anomaly. A closed form analytic solution is not known,

but the solution can be found through a variety of iterative procedures. Danby recommends a generalization of Halley's iterative method with quartic convergence, and a particular initial guess. (Watch out though, there is a typo in Danby's code on p. 167 which reduces its convergence to cubic. Also, Danby's convergence criterion is not strict enough.) For problems in which the orbits may become hyperbolic a solution of the initial value problem in terms of universal variables is probably preferred.

The integration of the system across a delta function is trivially accomplished in canonical Cartesian coordinates since in this system the interaction Hamiltonian depends only on the coordinates. The coordinates are unchanged since the interaction Hamiltonian does not depend on the momenta. The momenta each receive a kick proportional to the generalized force, which is derived in the usual way as minus the derivative of the interaction Hamiltonian with respect to the conjugate coordinate. The differentiation is straightforward and will not be presented here. We just mention a couple of key points. Note that with an appropriate rearrangement of terms the direct contributions to all of the disturbing accelerations can be evaluated in $o(n^2)$ operations, where n is the number of planets, and all of the indirect contributions can be evaluated in $o(n)$ operations. Also, some contributions to the forces arise as small differences of nearly equal quantities. The numerical inaccuracies which would be incurred by a straightforward evaluation of these expressions can be avoided by using the same trick used to avoid a similar difficulty encountered in Encke's method (see Danby, 1988). Closed form expressions for this trick are given by Potter (1962).

2.5 Refinement of the Mapping Method

There are some general refinements to the mapping method that can be made. We consider Hamiltonians of the general form

$$H = H_0 + H_1. \tag{2.22}$$

Both H_0 and H_1 may depend on *all* of the coordinates and momenta, though to make a mapping it is necessary that each part in the absence of the other part be integrable.

In this paper attention is focused on problems for which $H_0 >> H_1$. The basic idea behind the refinement of the mapping method is that rather than using a single delta function per mapping period, one mapping period can consist of a series of delta functions with possibly different amplitudes and various phases, with the amplitudes and phases chosen so that the mapping will have better properties. The property to optimize is left to our discretion. There are two obvious choices. One choice is to optimize the "order" of agreement of the Taylor series of the actual solution in time with one step of the mapping, treating the mapping as if it were a symplectic numerical integration algorithm. The other choice is to optimize the agreement of the mapping Hamiltonian with the true Hamiltonian, making the differences have as high a frequency as possible so that the average effect of the differences will be as small as possible. Curiously, the two choices are not equivalent. Since the averaging principle is at the core of our reasoning, the second choice, to make the map Hamiltonian agree as much as possible with the true Hamiltonian, will be considered first.

High frequency terms are added so that the corresponding mapping Hamiltonian has the form

$$H_{Map} = H_0 + \Phi(\Omega t) H_1, \tag{2.23}$$

where

$$\Phi(t) = 2\pi \sum_{i=0}^{N-1} a_i \delta_{2\pi}(t - 2\pi d_i). \tag{2.24}$$

There are N delta functions per mapping period, with amplitudes a_i, and phases d_i which are chosen in the interval $0 \leq d_i < 1$. Written as a Fourier series,

$$\Phi(t) = \sum_{i=0}^{N-1} a_i \sum_{n=-\infty}^{\infty} \cos(n(t - 2\pi d_i)) \tag{2.25}$$

$$= \sum_{n=-\infty}^{\infty} A_n \cos(nt) + \sum_{n=-\infty}^{\infty} B_n \sin(nt), \tag{2.26}$$

where

$$A_n = \sum_{i=0}^{N-1} a_i \cos(2\pi n d_i) \tag{2.27}$$

and

$$B_n = \sum_{i=0}^{N-1} a_i \sin(2\pi n d_i). \tag{2.28}$$

The coefficients A_n and B_n of each of the $\cos(nt)$ and $\sin(nt)$ terms provide constraint equations for the a_i and d_i. First, the average of $\Phi(t)$ over one mapping period must be unity for the average of the mapping Hamiltonian over a mapping period to equal the true Hamiltonian. The average is given by the $n = 0$ equation, which implies simply

$$A_0 = \sum_{i=0}^{N-1} a_i = 1. \tag{2.29}$$

For the two Hamiltonians to agree the coefficients of all the terms involving the mapping frequency must be zero. This gives the set of equations which determine the coefficients for each $n \neq 0$:

$$A_n = \sum_{i=0}^{N-1} a_i \cos(2\pi n d_i) = 0 \tag{2.30}$$

$$B_n = \sum_{i=0}^{N-1} a_i \sin(2\pi n d_i) = 0 \tag{2.31}$$

Note that if the two coefficient equations for some $n > 0$ are satisfied, the corresponding two coefficient equations for $n < 0$ are also automatically satisfied. We would like to satisfy as many of the coefficient equations as is possible, beginning with those of lowest frequency (smallest $|n|$). For a given N there are $2N$ constants to be determined. Thus, it is expected that these $2N$ constants can be chosen to satisfy at least $2N$ coefficient equations. There is only a single coefficient equation for $n = 0$. Thus it should be possible to make the coefficients up to $n = N - 1$ equal to zero, plus one of the two coefficients for $n = N$. Of course, just making one coefficient equal to zero is not very useful since other terms of that frequency will remain.

Consider the $N = 1$ map first. In this case there is only a single delta function. The coefficient $a_0 = 1$, and no useful constraint is placed on d_0. In the $N = 2$ map, it is easy to see that the two a_i must be equal, and consequently $a_i = 1/2$, and the d_i must differ by $1/2$. No useful constraint is placed on the absolute phase. For general N, a solution of the coefficient equations is that the $a_i = 1/N$ and the d_i are evenly spaced with separation $1/N$. No useful constraint is placed on the absolute phases by the coefficient equations, since the one remaining constant cannot simultaneously satisfy both of the next set of coefficient equations for $n = N$. It

can be easily shown that this solution does more than satisfy the $2N-1$ equations which determined the coefficients, but in fact it satisfies all coefficient equations for all n that are not multiples of N. It seems likely that this is the only solution to the coefficient equations, though we have not been able to demonstrate this. The solution just mentioned is very simple indeed, it consists of N evenly spaced delta functions, all of equal amplitude $1/N$. Actually, this is nothing but the original single periodic delta function with a smaller mapping period.

2.6 Higher Order Mappings

Another possible approach to refine the mapping method is to use the flexibility introduced through the $2N$ constants a_i and d_i in $\Phi(t)$ to match the Taylor series of a mapping step to that of the actual solution. The mapping could then be viewed as a symplectic integrator. To match the Taylor series by brute force to high order turns out to be a formidable task, even with the aid of computer algebra. There is however a more abstract representation of the generalized mapping method that makes the problem of the determination of the constants more tractable.

Consider in more detail the consequences of the mapping Hamiltonian

$$H_{Map} = H_0 + \Phi(\Omega t) H_1. \tag{2.32}$$

Again, both H_0 and H_1 may depend on all the coordinates and momenta. Between the times when the delta functions act the time evolution of the system is governed solely by H_0. On the other hand, the evolution across the delta functions is determined solely by H_1, even though the coordinates and momenta which are being affected may also appear in H_0. This may be seen through a simple limit process in which the delta function $2\pi\delta(\Omega t)$ is represented as the limit as $\Delta \to 0$ of a step function which is non-zero in the interval $0 < t < \Delta$ with constant magnitude $2\pi/\Omega\Delta$. Taylor expanding the solution across the delta function, including terms due to both H_0 and H_1, and then taking the limit of the result as Δ goes to zero, it is easily shown that the terms coming from H_0 do not contribute. In fact, it can also be seen that the

evolution of the system through the delta function is the same as if the system evolved solely according to H_1 for a time $\Delta t = 2\pi/\Omega$. The evolution through the generalized map with H_1 multiplied by a sequence of weighted delta functions of amplitudes a_i at times $d_i\Delta t$ can then be interpreted in the following way. First, the system evolves according to H_0 for a time $d_0\Delta t$, then according to H_1 for a time $a_0\Delta t$, then according to H_0 for a time $(d_1 - d_0)\Delta t$ up to the next delta function, then according to H_1 for a time $a_1\Delta t$, and so on until after the last delta function, whereupon the system evolves to the end of the mapping period according to H_0 for a time $(1 - d_{N-1})\Delta t$.

Now, it is well known that the Taylor series of a function can be written formally as an exponential

$$f(t_0 + \Delta t) = \exp(\Delta t \frac{d}{dt}) f(t)\Big|_{t=t_0}. \tag{2.33}$$

Also, the total time derivative of a function of the phase space coordinates (the n coordinates and n momenta for an n degree of freedom system), and possibly also the time, can be written in terms of a Poisson bracket with the Hamiltonian which governs the time evolution of the system

$$\frac{df}{dt} = [f, H] + \frac{\partial f}{\partial t}. \tag{2.34}$$

Let L_H represent the Poisson bracket operator $L_H f = [f, H]$, then the time evolution of a function which depends on time only through the phase space coordinates can be written

$$f(x,p)|_{t_0+\Delta t} = e^{\Delta t L_H} f(x,p)\Big|_{t=t_0}. \tag{2.35}$$

In particular, this is true for the individual coordinate functions

$$x|_{t_0+\Delta t} = e^{\Delta t L_H} x\Big|_{t=t_0}. \tag{2.36}$$

Furthermore, for a function which depends on time only through the phase space coordinates, the time evolved function must equal the function of the time evolved coordinates and momenta

$$e^{\Delta t L_H} f(x,p)\Big|_{t=t_0} = f(e^{\Delta t L_H} x, e^{\Delta t L_H} p)\Big|_{t=t_0}. \tag{2.37}$$

Of course these are nothing other than the basic equations governing Lie transformations (see e.g. Steinberg, 1988).

The operator that generates the Taylor series for the generalized map can then be formally written

$$e^{b_N \Delta t L_{H_0}} \cdots e^{a_1 \Delta t L_{H_1}} e^{b_1 \Delta t L_{H_0}} e^{a_0 \Delta t L_{H_1}} e^{b_0 \Delta t L_{H_0}}, \tag{2.38}$$

where the $b_i = d_i - d_{i-1}$, with $b_0 = d_0$ and $b_N = 1 - d_{N-1}$ Of course the exponentials cannot be simply combined since the operators L_{H_0} and L_{H_1} do not commute. (If the operators commuted the calculation of the time evolution would be trivial.) For time evolution which is governed by the true Hamiltonian the operator that generates Taylor series is just

$$e^{\Delta t L_H} = e^{\Delta t (L_{H_0} + L_{H_1})}. \tag{2.39}$$

Equating these two expressions up to some order in Δt gives constraint equations which must be satisfied by the a_i and b_i. The problem of determining the constants reduces to a problem in the algebra of exponentials of non-commuting operators. Though in principle some special property of L_{H_0} and L_{H_1} might be used to simplify the determination of the constants for a particular Hamiltonian, the solutions to date have treated them simply as general non-commuting operators. Replacing $\Delta t L_{H_0}$ by B and $\Delta t L_{H_1}$ by A, the coefficients are determined by requiring

$$e^{A+B} = e^{b_N B} \cdots e^{a_1 A} e^{b_1 B} e^{a_0 A} e^{b_0 B}, \tag{2.40}$$

be satisfied to a specified order in the products of the non-commuting operators A and B.

The solution of these equations to second order is simply achieved for $N = 1$: $b_0 = b_1 = 1/2$, and $a_0 = 1$. The result is the "generalized leap-frog": a half step following H_0, followed by successive whole steps alternately following H_1 and H_0, ending with a half step of H_0. Note that in this case, the delta functions are all equally spaced in time. Thus the agreement of the mapping Hamiltonian with the true Hamiltonian, as discussed in the previous section, is maximally retained. The previously noted extra freedom of the phase of the evenly spaced delta functions is being used to make the map accurate to second order, at no extra cost.

The solution of these equations to fourth order is possible with $N = 3$. This has been independently accomplished by Forest and Ruth (1990) and Yoshida (1990), also using Eq.(2.40). The solution is the same as that found in a more restricted formalism by Candy and Rozmus (1990) and earlier by Neri (1988). Yoshida (1990) has found sixth and eighth order solutions for $N = 7$ and $N = 15$ respectively. Forest (1990) has also obtained sixth order solutions. These authors were interested in extending the symplectic integration method of Ruth (1983) to higher order. Our formulation of the coefficient equations in terms of exponentials of non-commuting operators had a different motivation and was carried out before we became aware of the work of Forest and Ruth (1990), and Yoshida (1990).

It is interesting to note that in all the known solutions of the coefficient equations some of the b_i are negative, except the first and second order methods (the generalized leap-frog). It appears that to get higher order it is necessary to take some steps backward in time. However, even though the coefficients solve the coefficient equations, these negative steps cannot simply be represented in the scalar Hamiltonian (2.32) from which the coefficient equations were derived, since in the Hamiltonian only the time at which the delta functions act is specified and not any additional order of application. A Hamiltonian could be written for the higher order maps as

$$H_{Map} = \Phi_0(\Omega t)H_0 + \Phi_1(\Omega t)H_1. \tag{2.41}$$

where $\Phi_0(t)$ and $\Phi_1(t)$ are similar to $\Phi(t)$, but with different coefficients. Now between the delta functions the Hamiltonian is zero and there is no evolution. The higher order maps can be generated if the delta functions for $\Phi_0(t)$ and $\Phi_1(t)$ are interleaved with the proper amplitudes. For Hamiltonians which have been separated into two parts H_0 and H_1 of comparable magnitude, this seems to be a satisfactory solution, though it is not clear that anything is gained by representing the maps in this manner. On the other hand, for Hamiltonians for which H_0 is much larger than H_1 a mapping Hamiltonian which introduces high frequencies proportional to H_0 seems like a bad idea. The first corrections to the evolution as deduced by, for example, Von Zeipel transformations will be large. Perhaps there is better way to represent the

Hamiltonian for the higher order methods. An interesting possibility is to search for higher order methods using Hamiltonian (2.32), but with the added constraint that all the b_i be positive. We have attempted to find two and three delta function maps of third order with all positive b_i, but were unsuccessful. It is important to determine whether or not such solutions exist in general.

At present then, there are two options available: (1) If the existence of an explicit Hamiltonian is desired, and the agreement of the map Hamiltonian with the true Hamiltonian is to be maximized, then the best solution is to take all the delta functions to be of equal amplitude and evenly spaced. Without any degradation of the agreement of the two Hamiltonians, the map can be made second order by choosing the phase of the delta functions so that the calculation begins and ends with a half step of H_0. (2) If the order of the mapping is to be increased beyond second order, any of a number of known solutions of the coefficient equations can be used. Apparently though, higher order comes at the cost of losing, to some extent, our original motivation which connects a mapping Hamiltonian via the averaging principle to the true Hamiltonian, particularly for a system for which $H_0 >> H_1$. Nevertheless, either of these alternatives can be used with the n-body mapping components described here.

The n-body mapping is then a composition of individual Kepler steps for each of the planets with kicks derived from the interaction Hamiltonian, each appropriately weighted to form either the generalized leap-frog version of the map or a higher order version of the map. Note that since each of the Keplerian orbits is advanced independently of all the others, they can be advanced in parallel.

2.7 Other Symplectic Integrators

It is instructive to compare the generalized maps described here with other symplectic integrators. The symplectic integration scheme of Ruth (1983) (see also de Vogelaere, 1956) is based on a time step given in terms of a mixed variable generating function,

$$F(x, p') = xp' + S(x, p'), \tag{2.42}$$

where the primed variables are the new variables after the time step. Nothing here depends on the dimension of the phase space, so for simplicity we use a single degree of freedom. Choosing $S(x,p') = \Delta t H(x,p')$, this generating function gives the canonical transformation of variables

$$p = \frac{\partial F}{\partial x} = p' + \Delta t \frac{\partial H(x,p')}{\partial x} \tag{2.43}$$

$$x' = \frac{\partial F}{\partial p'} = x + \Delta t \frac{\partial H(x,p')}{\partial p'}. \tag{2.44}$$

This step is canonical and only approximates the evolution of the system under $H(x,p)$ to first order in Δt. Higher order generating functions have been derived by Ruth(1983), Menyuk(1984), and Channel and Scovel (1988). In general, the transformation from x and p to x' and p' is only given implicitly, since it is based on a mixed variable generating function. In the special case where the Hamiltonian can be written in the form

$$H(x,p) = T(p) + V(x) \tag{2.45}$$

the transformation can be written explicitly:

$$p' = p - \Delta t \frac{\partial V(x)}{\partial x} \tag{2.46}$$

followed by

$$x' = x + \Delta t \frac{\partial T(p')}{\partial p'} \tag{2.47}$$

Ruth (1983) found higher order symplectic integrators both by using higher order generating functions for the step, and by composing low order steps and adjusting constants to achieve higher order. The latter method is followed by Neri (1988) and Candy and Rozmus (1990) for Hamiltonians of the form of Hamiltonian (2.45). Forest and Ruth (1990) also describe their method in terms of a composition of steps achieved through a mixed variable generating function.

The mapping method has been described in terms of the averaging method which originally motivated it, and implemented in terms of mapping Hamiltonians containing periodic sequences of delta functions. This is quite different from the generating

function description of the symplectic integrators following Ruth (1983). Nevertheless, mappings derived along the lines Wisdom (1982) are closely related to the symplectic integration methods derived from Ruth (1983). In fact, for Hamiltonians of the form of Hamiltonian (2.45), they are identical. Consider a simple mapping Hamiltonian for Hamiltonian (2.45)

$$H_{Map} = T(p) + 2\pi\delta_{2\pi}(\Omega t)V(q). \tag{2.48}$$

Integrating across the delta function gives

$$p' = p - \Delta t \frac{\partial V(x)}{\partial x}, \tag{2.49}$$

with $\Delta t = 2\pi/\Omega$. Then integrating between the delta functions gives

$$x' = x + \Delta t \frac{\partial T(p')}{\partial p'}. \tag{2.50}$$

These are the same equations as those obtained with the mixed variable generating function. However, in the more general case in which the Ruth method will require the solution of implicit equations the mapping method presented here will give explicit equations. The two methods are then not equivalent.

It is clear that the mapping method based on periodic delta functions can also be thought of as a method of symplectic integration. It is sometimes the same as that derived from a generating function and sometimes distinct. In a sense, though, the mapping based on delta functions is more clearly related to the original system than is the symplectic integration step based on generating functions. In the Ruth school, the basic idea is that a finite difference method that is exactly canonical may be subject to fewer artifacts than a finite difference method that is not canonical. For instance, a non-canonical finite difference scheme may have attractors, in contradiction to the well known fact that Hamiltonian systems do not have attractors. The finite difference scheme then has a possible behavior that cannot belong to the real system. The Ruth integrators are symplectic as the actual system they are modeling is symplectic, and for small enough time-step the error in the step becomes arbitrarily small, but whether the symplectic integrator should give a good approximation to the long term

dynamics is not clear. This point was reiterated by Zhong and Marsden (1988). One method of deriving these symplectic integrators (see Forest and Ruth, 1990) is to make a chain of canonical transformations as described above and then require that the Hamiltonian, when expressed in terms of the new variables, matches the actual Hamiltonian to some order. Since the match is not perfect, each individual step of a symplectic integrator makes some error in the Hamiltonian. There is no guarantee, nor any reason to expect, that the repeated composition of such steps, will not lead further and further from the true Hamiltonian, even though the composition is canonical. However, this catastrophy does not seem to happen, though the reason remains unclear. On the other hand, the mappings derived from Wisdom (1982) are explicitly derived from a Hamiltonian. The differences between the true Hamiltonian and the mapping Hamiltonian are arguably unimportant to the long term behavior on the basis of the averaging principle. Furthermore the differences are explicit. In some cases it is possible to derive a correction from mapping variables to true variables by eliminating the extraneous high frequency terms by Von Zeipel transformations (e.g. Tittemore and Wisdom 1988). The basic Ruth step is canonical, but it is given in terms of a generating function, not as the time evolution of a Hamiltonian (though Menyuk, 1984, calls the generating function a "discrete Hamiltonian"). It is not clear to us whether or not a system Hamiltonian can be written for the implicit Ruth maps.

2.8 Simpler N-Body Maps

There is a simpler way to get symplectic maps for the n-body problem. Note that the basic n-body Hamiltonian (2.6) is in the form of Hamiltonian (2.45). Thus a particularly simple map for the n-body problem can be obtained by letting H_0 be the kinetic energy and H_1 be the potential energy. Then any of the mapping Hamiltonians we have discussed can be used. In particular, high order mappings can be made with Hamiltonian (2.41). The second order form of the map, the generalized leap frog, becomes in this case simply the ordinary leap frog integrator.

There are two obvious disadvantages of these simple n-body maps. The most

severe is that the basic Keplerian motion of the orbits is not taken into account. Thus the number of steps must be large enough to stably and accurately negotiate each passage of the planet around the sun. A second disadvantage is that the motion of the central mass must also be integrated. It should be noted that Cowell's form of the equations of motion, which are written in terms of simple relative coordinates with respect to the central mass (see Brouwer and Clemence, 1961), are not in Hamiltonian form. A map naively based on them would not have the desired property of being symplectic.

Higher order maps of this simple form have already been applied to the n-body problem by Gladman and Duncan (1990) and Kinoshita, Yoshida, and Nakai (1990). Note that in both of these applications the mappings are used in what may be called the "qualitative" mode of operation. The step size is relatively large and the truncation error is much larger than the machine precision. Of course, this may not matter for qualitative investigations, since the truncation error does not seem to accumulate in a bad way for these symplectic maps, and the energy error is observed to oscillate, at least locally. Gladman and Duncan take 300 steps per orbit; Kinoshita, et al. take about 600 steps per orbit. The relative error in the value of the Hamiltonian is of order 10^{-9} for Kinoshita, et al., and somewhat larger in the other study. In both cases the error is much larger than the available precision of the machine and is obviously dominated by truncation error. The simple fourth order symplectic integrators used in these papers evaluate the accelerations three times per step. Thus they require about 2000 function evaluations per orbit to achieve a local relative energy error of 10^{-9}. These errors should be compared with the errors in a conventional numerical integration. For example, for the twelfth order Stormer predictor, which was used in the Digital Orrery calculations, the truncation error is of order the machine precision with around 100 steps per orbit. The Stormer predictor evaluates the accelerations only once per step. Even after nearly a billion years the relative energy error in the Digital Orrery integration was only of order 10^{-10}, and locally the variation of the relative energy is much smaller. Even used in the qualitative mode of operation the simple symplectic integrators do not appear to be competitive with traditional

integrators.

Consider the use of the simple n-body maps in a "high accuracy" mode of operation in which the step size is chosen to be small enough so that the truncation error is of order the machine precision. Assuming the truncation error is proportional to the fifth power of the step size for these fourth order methods, for the relative energy error to reach machine precision (which we take to be about 10^{-16}) requires about 45,000 function evaluations per orbit. Of course, higher order methods need to take significantly fewer steps per orbit. Suppose the relative energy error from truncation can be written $\Delta = C(h/N)^{o+1}$, where h is the step-size divided by the orbital period, o is the order, N is the number of function evaluations per step, and C is an error constant. We presume that the error constant in this form is comparable for all of the higher order methods; the factor of N is just a guess, and works in favor of the higher order methods. Using this estimate we find that even for the eighth order method of Yoshida (with 15 function evaluations per step), achieving a relative energy error of order the machine precision requires 400 function evaluations per orbit. Thus even the high order versions of the simple n-body maps may still be inefficient compared to traditional high accuracy integrators. Of course, the relative inefficiency may be outweighed by a better long-term growth of error. To our knowledge the long-term growth of error for the simple symplectic n-body integrators has not yet been carefully examined, particularly in the "high accuracy" mode of operation where the truncation error is of order the machine precision.

Consider in the same manner the possibility of using the n-body maps described in this paper in the "high accuracy" mode. From a numerical integration point of view, the basic difference between these methods and the simple methods just described is that the error constant in these new maps may be expected to be smaller by about the ratio of the planetary masses to the central mass, μ. For our solar system μ is about 10^{-3} . The number of steps per orbit required to achieve the same truncation error as the simple maps is smaller by a factor of $\mu^{-1/(o+1)}$. For a fourth order method with $\mu = 10^{-3}$ this factor is only about 4. For the eighth order method it is about 2. Considering the fact that the steps in the Kepler-based n-body maps are a little more

expensive than those in the simple n-body maps, it is not clear that any advantage is gained by using the maps presented here over the simpler maps, at least in the "high accuracy" mode of operation. However, there may be an advantage to using the Kepler-based maps for orbits with high eccentricity. In this case, the simple n-body maps must take many more steps per orbit to stably and accurately execute the orbit, since the basic Kepler motion must be integrated as well. On the other hand the n-body maps presented in this paper exactly represent a pure Kepler orbit at any eccentricity. Tests in the circular and elliptic restricted problems indicate the Kepler based n-body maps suffer no significant loss of stability or accuracy at high eccentricity. In this case there may be a significant advantage in using them over the simple maps even in the "high accuracy" mode.

On the other hand, consider the use of the n-body maps introduced in this paper in the "qualitative" mode of operation. Typically, efficient traditional integrators take on the order of 100 steps per orbit. We have found that in solar system integrations the qualitative behavior is reliably reproduced with as few as ten steps per orbit. Such a small number of steps per orbit is stable here because the Kepler motion is represented exactly and does not have to be rediscovered each orbit. The reduction in the number of function evaluations by a factor of 10 accounts roughly for the order of magnitude greater speed of the new mapping method over traditional integrators. The new n-body maps are the clear winners for qualitative studies.

Of course, the relative merits of the various methods in the two different modes of operation should be studied more thoroughly to check the estimates given here.

2.9 The Outer Planets for a Billion Years

We have carried out numerous tests of the new n-body maps. First, a number of surfaces of section for the circular restricted three-body problem were computed with the new map and compared to sections computed with the conventional Bulirsch-Stoer numerical integration algorithm. The agreement was excellent and provided valuable initial experience with the new maps. These tests demonstrated the reliabil-

ity and efficiency of the map at high eccentricity. The n-body maps have also been implemented for the planar elliptic restricted three-body problem. The numerical integrations reported in Wisdom (1983), which also used the conventional Bulirsch-Stoer algorithm, were all repeated with the map, with particular attention to whether the map would give the correct diagnosis of whether the trajectory was chaotic or quasiperiodic. In every case, the map agreed with the earlier results. Of course, the jumps in eccentricity were also recovered. Note that the codes for the various versions of the restricted three-body problem can be written to take advantage of the known fixed orbit of the two massive bodies. Rather than present these initial tests in detail, we present a much more stringent test. We have used the map to compute the evolution of the outer planets, including Pluto as a test particle, for about 1.1 billion years. For this problem the evolution has already been computed for 845 million years using conventional integration techniques on the Digital Orrery (Sussman and Wisdom, 1988), and comparison can be made to those results.

We have chosen to use the second order version of the mapping, which optimizes the agreement of the mapping Hamiltonian with the true Hamiltonian in accordance with our original motivation based on the averaging principle. We have used the exact form of the interaction Hamiltonian, and Pluto is given a Jacobi index below those of the massive planets. Of course, in order to make comparisons the initial conditions and parameters must be the same as those used in the Digital Orrery integrations (Applegate, et al. 1986). The only parameter left to choose is the step size, or mapping period. The map is used in the "qualitative" mode and the step-size is chosen to be relatively large. A number of preliminary tests indicate that the map does not work well for this problem if fewer than five steps are taken per Jupiter orbit period, which is about 12 years. To add a margin of safety, a step size of 1 year was chosen. This may be compared to typical steps of 40 days or less that have been used in other studies of the outer planets using conventional numerical integration techniques. The relative energy error oscillates as expected, and, using this step-size, has a rather large peak to peak amplitude of about 10^{-5}. The map is remarkably fast. A billion year evolution of the outer planets takes only 14 days on a Hewlett-Packard

HP9000/835 RISC workstation.

All of the principal results of Sussman and Wisdom (1988) are reproduced in the mapping evolution. For example, the argument of perihelion of Pluto again displays a 34 million year modulation. The quantity $h = e \sin \varpi$, where ϖ is the longitude of perihelion, displays its strong 137 million year period. This is illustrated in Fig. (2-1) which is to be compared with h as computed using the Stormer multistep predictor on the Digital Orrery, shown in Fig. (2-2). The two plots are not identical, but the similarity is astounding. The inclination of Pluto again displays even longer periods. Fig. (2-3) presents the inclination of Pluto for a billion years as computed with the map; and, for comparison, Fig. (2-4) presents the inclination of Pluto from the Digital Orrery computation. In the region of overlap the two plots are again remarkably similar. The observed differences can probably all be attributed to the different sampling times in the latter part of the Digital Orrery computation, and to slightly different frequencies in the two evolutions. Even used in the "qualitative" mode these n-body maps are remarkably good. It is interesting that at the end of the mapping calculation, which was longer than the Orrery integration, the inclination reaches a new maximum which gives the impression of a secular increase and at least indicates the presence of periods longer than a billion years in the motion of Pluto. The motion of Pluto appears to be inexhaustible, a property which is consistent with the numerical evidence that the motion of Pluto is chaotic.

Finally, the chaotic character of the motion of Pluto has been confirmed using the mapping method. The divergence of nearby trajectories has been computed using the two-trajectory method. The plots are qualitatively the same as the one shown in Sussman and Wisdom (1988). In that study the final Lyapunov exponent calculation was begun halfway through the computation; here the Lyapunov exponent calculation was started at the beginning of the run. Since the two calculations are not directly comparable, they will not be shown here. The timescale for exponential divergence is again about 20 million years. This confirmation of the positive Lyapunov exponent on a different computer with a different length floating point mantissa, with such a different integration method, using a somewhat arbitrarily chosen step-size which is

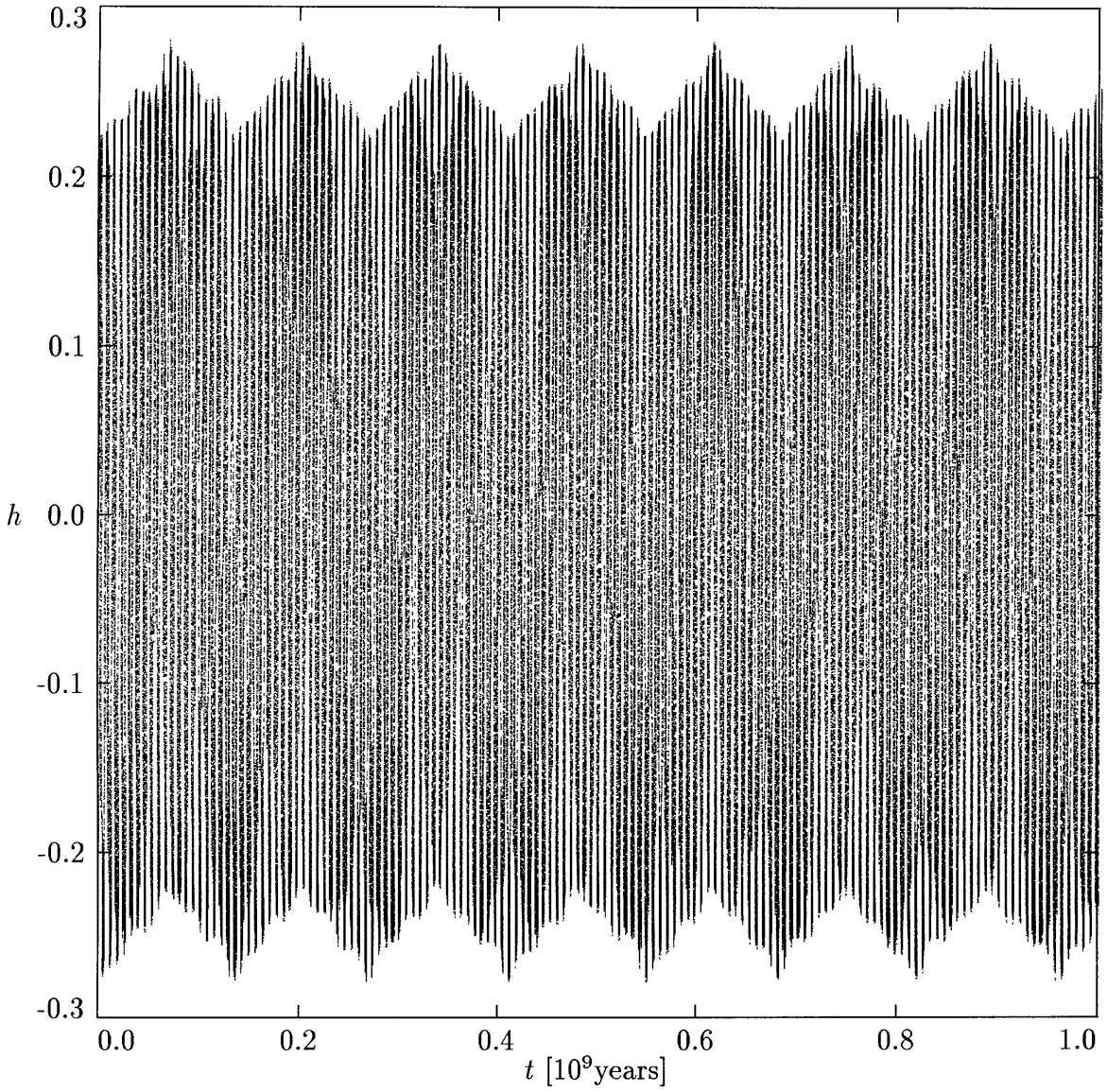


Figure 2-1: The orbital element $h = e \sin \varpi$ for Pluto as computed with the mapping.

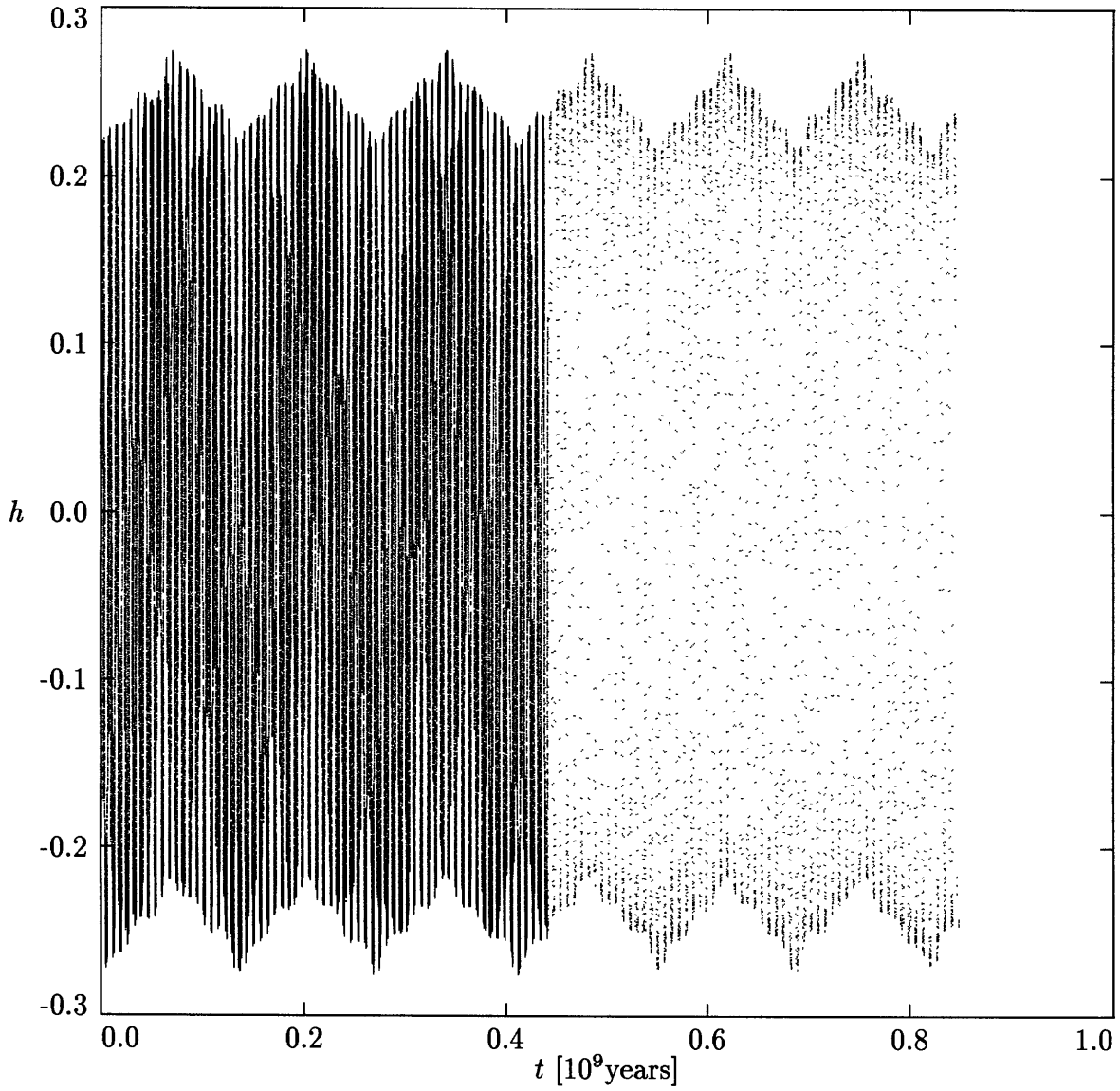


Figure 2-2: The orbital element $h = e \sin \varpi$ for Pluto as computed with a conventional numerical integration method on the Digital Orrery. The sampling frequency was decreased in the latter part of the run.

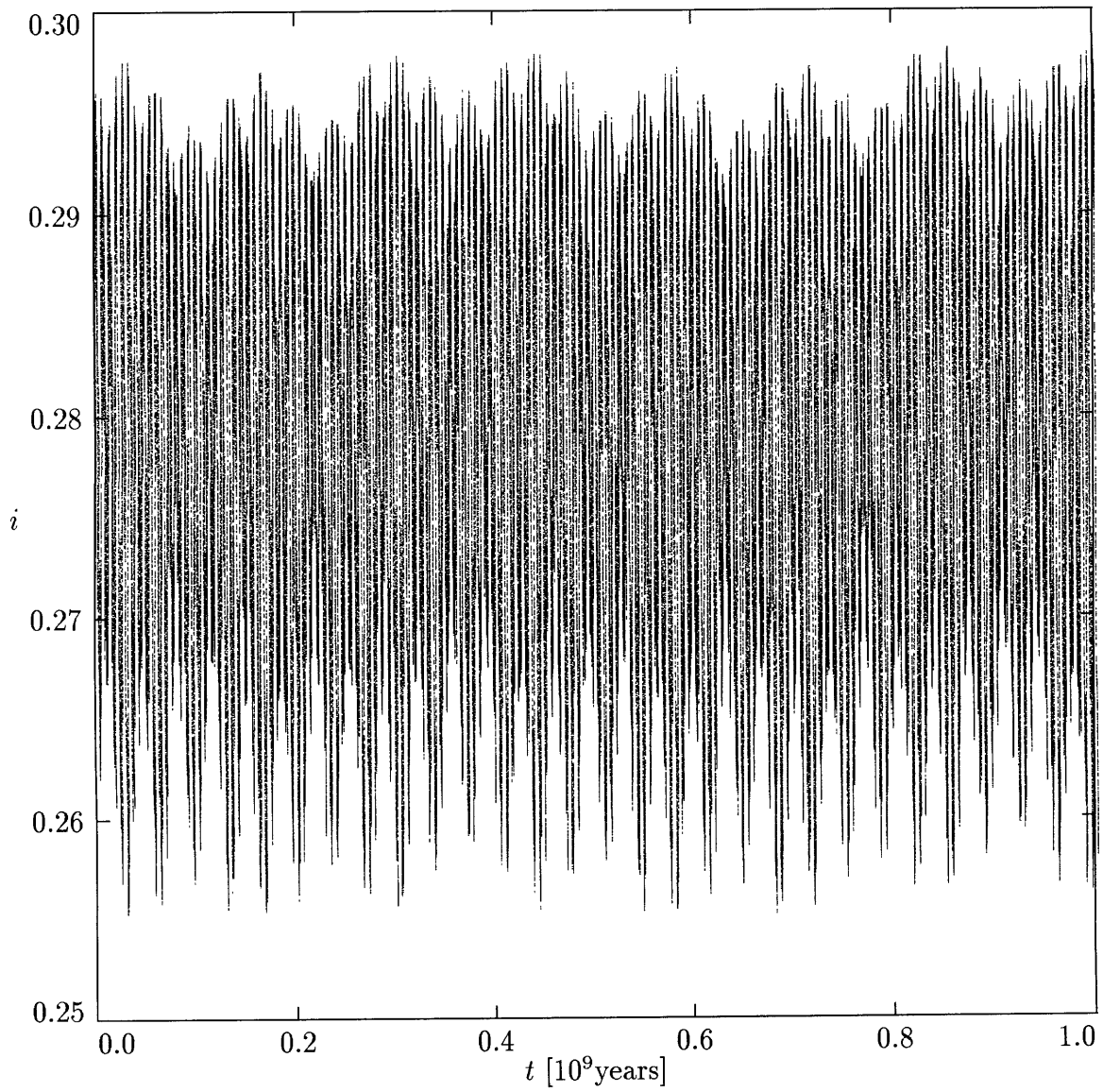


Figure 2-3: The inclination of Pluto (in radians) for a billion years as computed with the mapping.

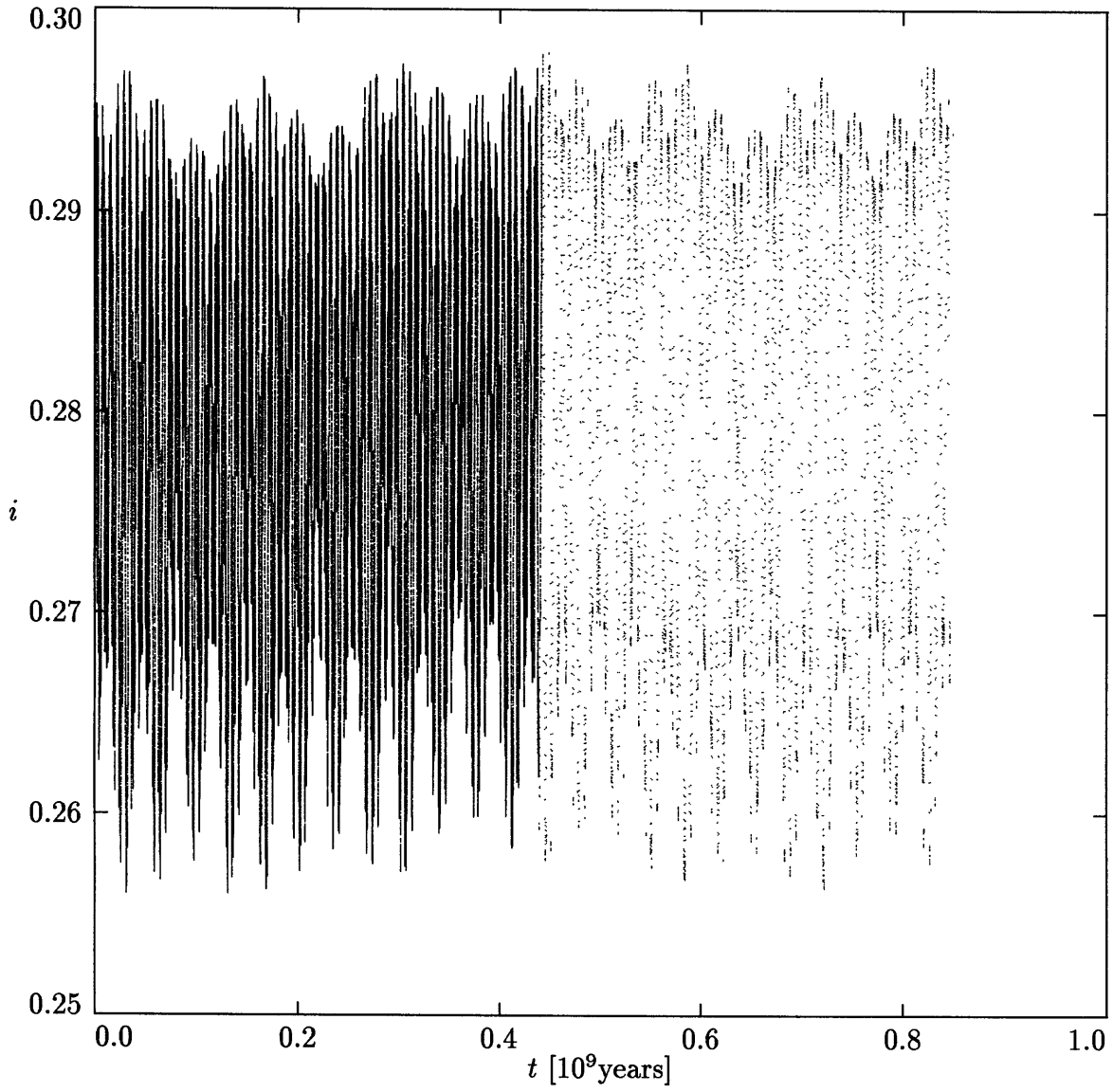


Figure 2-4: The inclination of Pluto (in radians) for 845 million years as computed with a conventional numerical integration method on the Digital Orrery.

an order of magnitude larger than the special step-size used in the Digital Orrery integrations, considerably strengthens the conclusion of Sussman and Wisdom (1988) that the motion of Pluto is chaotic. Of course, it is probably wise to remain a little suspicious until the dynamical mechanisms are properly understood.

This is now the longest evolution of the outer planets to date. The remarkable agreement of this evolution with that computed with the Digital Orrery is a strong testimony to the validity of the averaging principle, and to the usefulness of the new n-body maps.

2.10 References

Applegate, J.H., Douglas, M.R., Gursel, Y., Sussman, G.J., and Wisdom, J. (1986). Astron. J. **92**, 176.

Arnold, V.I. (1974). *Mathematical Methods of Classical Mechanics* (Springer-Verlag, New York).

Brouwer and Clemence, (1961). *Methods of Celestial Mechanics* (Academic Press, New York).

Candy and Rozmus (1990). preprint.

Channell and Scovel (1988). Los Alamos National Laboratory, LA-UR-88-1828.

Chirikov, B.V. (1979). Phys. Rep. **52**, 263.

Cohen, C.J., Hubbard, E.C., and Oesterwinter, C. (1973), Astron. Papers Amer. Ephemeris **22**, 1.

Danby, J.M.A. (1988). *Fundamentals of Celestial Mechanics*, (Willmann-Bell, Richmond).

DeVogelaere, R. (1956). University of Notre Dame Report 4.

Forest, E. (1990). preprint.

Forest, E. and Ruth, R.D. (1990) Physica **D 43**, 105.

Gladman, B. and Duncan, M. (1990). preprint.

Kinoshita, H. and Nakai, H. (1984) Celest. Mech. **34**, 203.

Kinoshita, H., Yoshida, H., Nakai, H. (1990). in *Proceedings of the Twenty-Third Symposium on "Celestial Mechanics."*, Eds. H. Kinoshita and H. Yoshida, (Kyoto), p. 1.

Laskar, J. (1989). Nature **338**, 237.

Laskar, J. (1990). preprint.

Malhotra, R. (1990). Icarus **87**, 249.

Malhotra, R. and Dermott, S., (1990). Icarus **85**, 444.

Menyuk, C.R. (1984). Physica **11D**, 109.

Murray, C.D. (1986). Icarus **65**, 70.

Murray, C.D. and Fox, K. (1984). Icarus **59**, 221.

Neri, F. (1988). preprint.

Plummer, H.C. (1960). *An Introductory Treatise on Dynamical Astronomy*, (Dover, New York).

Quinn, T.R., Tremaine, S., and Duncan, M. (1991) preprint.

Richardson, D.L. and Walker, C.F. (1987) in *Astrodynamics 1987* (Advances in the Astronautical Sciences, Vol. 65), Soldner, J.K., Misra, A.K., Lindberg, R.E., and Williamson, W., eds. (San Diego, Univelt), 1473.

Roy, A.E., Walker, I.W., Macdonald, A.J., Williams, I.P., Fox, K., Murray, C.D., Milani, A., Nobili, A.M., Message, P.J., Sinclair, A.T., and Carpino, M. (1988). Vistas in Astronomy **32**, 95.

Ruth, R.D. (1983). IEEE Trans. Nuclear Science **NS-30**, 2669.

Sidlichovsky, M. and Melendo, B. (1986). Bull. Astron. Inst. Czechosl. **37**, 65.

Steinberg, S. (1988). in *Lie Methods in Optics*, Eds. Mondragon and Wolf, (Springer-Verlag, New York), p. 45.

Sussman, G.J. and Wisdom, J. (1988). Science **241**, 433.

Tittemore, W.C. (1990). Science, **250**, 263.

Tittemore, W.C. and Wisdom, J. (1988). Icarus **74**, 172.

Tittemore, W.C. and Wisdom, J. (1989). Icarus **78**, 63.

Tittemore, W.C. and Wisdom, J. (1990). Icarus **85**, 394.

Wetherill, G. (1985). Meteoritics, **18**, 1.

Wisdom, J. (1982). Astron. J. **87**, 577.

Wisdom, J. (1983). Icarus **56**, 51.

Wisdom, J. (1985a). Nature **315**, 731.

Wisdom, J. (1985b). Icarus **63**, 272.

Wisdom, J. (1987). Icarus **72**, 241.

Yoshida, H. (1990). Phys. Lett. **A 150**, 262.

Zhong, G. and Marsden, J.E. (1988). Phys. Lett. **A 133**, 134.

Chapter 3

Symplectic Maps for the N-Body Problem: Stability Analysis [2]

3.1 Introduction

We recently introduced a new symplectic mapping method for studying the long term evolution of n-body problems with a dominant central mass (Wisdom and Holman, 1991, hereafter WH91). The method shows promise of being a valuable tool in the numerical exploration of planetary and satellite n-body systems. Tests of the method in various problems have indicated that the new mapping method can be an order of magnitude faster than other methods of numerical integration. It has already been used to carry out record-breaking long-term integrations of the solar system. In particular, Sussman and Wisdom (1992) used the mapping method to integrate the whole solar system for 100 million years. This integration confirmed the result of Laskar (1990) that the evolution of the solar system is chaotic with a surprisingly short timescale for exponential divergence of only 4 million years. In this paper, we examine more carefully the dynamical mechanisms which govern the stability of the mapping method. Our goal is to clarify the regime of applicability and understand more clearly the limitations of the mapping method.

We present here a novel technique for analyzing the non-linear stability of a nu-

[2] Adapted from Wisdom, J. and M. Holman (1992). Symplectic Maps for the N-body Problem: Stability Analysis. *Astron. J.* **104**, 2022–2029.

merical integration technique. The mapping is derived as the time evolution of a Hamiltonian. We analyze the mapping Hamiltonian as we would any other dynamical system using the tools of non-linear dynamics. The true Hamiltonian and the mapping Hamiltonian differ by the addition of a suite of resonances associated with the mapping stepsize. The evolution computed with the mapping approximates the true evolution provided these stepsize resonances do not significantly affect the evolution. We identify the principal stepsize resonances and analyze each in detail. Gross instability of the mapping method is associated with overlap of the principal stepsize resonances.

3.2 Mapping Method

The mapping method is a generalization of the resonance mapping method of Wisdom (1982, 1983). The derivation of the mapping method is detailed in WH91. Briefly, the Hamiltonian for the n-body problem can be written

$$H = H_{Kepler} + H_{Interaction}, \tag{3.1}$$

where H_{Kepler} represents the basic Keplerian motion of each of the planets around the dominant central mass, and $H_{Interaction}$ represents the interactions among them. Elimination of the center of mass may be accomplished by using Jacobi coordinates or canonical heliocentric coordinates. There is considerable freedom in deriving maps; see WH91 for details. The mapping method is based on the averaging principle: rapidly varying terms do not contribute significantly to the long-term evolution. Thus rapidly varying terms can be added or subtracted from the Hamiltonian with impunity. The simplest mapping is obtained by introducing extra high-frequency terms through periodic delta functions

$$H_{Map} = H_{Kepler} + 2\pi\delta_{2\pi}(\Omega t)H_{Interaction}, \tag{3.2}$$

where $\delta_{2\pi}(t)$ is a periodic sequence of Dirac delta functions with period 2π, and Ω is the mapping frequency. The time between delta functions is the stepsize $h = 2\pi/\Omega$.

The delta functions have the Fourier representation

$$\delta_{2\pi}(t) = \frac{1}{2\pi} \sum_{l=-\infty}^{\infty} \cos(lt). \tag{3.3}$$

Multiplying the interaction part of the Hamiltonian by the periodic delta function gives the original interaction Hamiltonian plus the same terms multiplied by terms of high frequency. The average of the mapping Hamiltonian over a mapping period gives the original n-body Hamiltonian. The advantage of introducing the delta functions is that the mapping Hamiltonian is locally integrable: between the delta functions each planet evolves along an unperturbed Keplerian orbit, and, also, the system is trivially integrated across the delta functions since the interaction Hamiltonian can be written solely in terms of the Cartesian coordinates. For the map to be a viable numerical method it is essential to be able to rapidly advance Keplerian orbits, and minimize intermediate canonical transformations. We eliminate the need for intermediate transformations by using Cartesian coordinates throughout; Keplerian orbits can be rapidly advanced directly in Cartesian coordinates using Gauss' f and g functions. Several extensions of the mapping method are presented in WH91.

3.3 Overview of Nonlinear Stability

Traditionally, the stability of an integrator is analyzed by applying the integrator to a linear system, and then solving the resulting set of linear difference equations. The linear stability of the method is then hoped to be relevant to the stability of the method when is it applied to the integration of non-linear problems. There is really no other choice, because the system of non-linear difference equations obtained in the approximation of a non-integrable dynamical system are too complicated to tackle directly. The Hamiltonian nature of the mapping offers a tremendous advantage for stability analysis. It allows us to study the non-linear stability of our mapping method using the usual tools of Hamiltonian dynamics. For background on Hamiltonian dynamics see, for example, Chirikov (1979) and Lichtenberg and Lieberman (1983).

An analogy will help clarify the following stability analysis. Let us construct a

simple mapping for the mathematical pendulum

$$H = \frac{1}{2}p^2 + \epsilon \cos\theta \tag{3.4}$$

According to our usual procedure we introduce a periodic sequence of delta functions into the "perturbation" part of the Hamiltonian (ignoring and, in fact, destroying the integrability of the original system)

$$H_{Map} = \frac{1}{2}p^2 + 2\pi\delta_{2\pi}(\Omega t)\epsilon \cos\theta. \tag{3.5}$$

This Hamiltonian is locally integrable: we easily compute the momentum change as the system crosses a delta function, and between the delta functions the angle rotates uniformly. The resulting map is

$$p' = p + \frac{2\pi}{\Omega}\epsilon \sin\theta \tag{3.6}$$

$$\theta' = \theta + p'\frac{2\pi}{\Omega}. \tag{3.7}$$

Higher order versions of this map could easily be constructed, as described in WH91. This map is a simple finite difference approximation to the evolution of the pendulum. As a consequence of being derived from a Hamiltonian, this finite difference scheme is symplectic, which in this case simply means it is area preserving. So we have a first order symplectic integrator for the pendulum. The stepsize, or mapping step, is $h = 2\pi/\Omega$. In terms of the scaled momentum $I = ph$, and the parameter $K = h^2\epsilon$, the map takes the form

$$I' = I + K \sin\theta \tag{3.8}$$

$$\theta' = \theta + I'. \tag{3.9}$$

We furthermore note that the dynamics is unchanged if we write the map in terms of a shifted momentum $J = I + 2\pi n$, where n is some integer. Thus, without any loss of dynamical possibilities we can consider the momentum to be periodically wrapped through the interval 0 to 2π, i.e. the momentum can be considered to be an angle variable. This is now the usual standard map (Chirikov, 1979). In the context of this paper though it is clearer to not restrict the momentum to be an angle. This

identification of the standard map as a symplectic finite difference approximation to the pendulum is well known (see, for example, Sanz-Serna and Vadillo, 1986).

The behavior of the standard map has been the object of a tremendous number of investigations, both mathematical and numerical. In the present context the review of Chirikov (1979) is most relevant. Though much is known about its behavior, it is not likely ever to be exhaustively understood. In what sense then does the standard map approximate the pendulum? Let us review the gross behavior of the standard map from this point of view.

The standard map exhibits a gross transition to largescale chaos roughly for $K > 1$. In that regime the dynamics has little to do with the pendulum. For small values of the parameter K (equivalently small mapping stepsize h) the phase space of the standard map displays a single pendulum-like resonance centered at $I = p = 0$, as well as periodic copies of this resonance, as described above, spaced by $\Delta I = 2\pi$ (equivalently $\Delta p = \Omega$). As long as we avoid the phase space near the periodic copies of our basic pendulum resonance the most notable departure of the mapping approximation to the pendulum from the pendulum itself is the appearance of a chaotic zone near the pendulum separatrix. The size of this chaotic zone has been estimated by Chirikov using the resonance overlap criterion. The resonances which overlap are the resonances between the libration period of the pendulum and the mapping period. There is an accumulation of these resonances as the unstable equilibrium is approached because the libration period of the pendulum diverges and as it diverges it successively matches every multiple of the mapping period. For our mapping approximation to the pendulum, application of the Chirikov formula for the width of the separatrix yields

$$\Delta\theta = \frac{8\pi^2}{h^{3/2}\epsilon^{3/4}} \exp\left(-\frac{\pi^2}{2h\epsilon^{1/2}}\right), \tag{3.10}$$

where $\Delta\theta$ is the half-width in θ of the chaotic zone near the unstable equilibrium. For small stepsizes h the size of this unwanted chaotic zone decreases very rapidly, as $\exp(-c/h)$. For example, even if we take only 10 mapping steps per small amplitude libration period of the pendulum (the natural timescale), the unwanted chaotic zone

has a fractional width of only about one percent. With 50 steps per libration period the size of the chaotic zone is reduced to less than one part in 10^{14}. If we are not interested in this particular region of the phase space, it is easy to avoid. In general, we are interested in some particular finite amplitude oscillation of the pendulum. The mapping will become grossly unstable for this particular trajectory if a stepsize h is chosen for which the chaotic zone near the separatrix is so large that the trajectory is engulfed in it. This clearly depends on the particular trajectory of interest; the closer the trajectory is to the separatrix the smaller the stepsize has to be to avoid this gross instability.

There are other artifacts introduced by the mapping approximation. From the Poincare-Birkhoff theorem we expect there are an infinite set of islands in the phase space of the standard map. Whenever the libration frequency is commensurate with the mapping frequency a chain of alternating stable and unstable periodic orbits appears in the phase space. Since the rationals are dense, the secondary islands associated with the stable periodic orbits are dense in the phase space. Of course these islands do not exist in the original pendulum problem. These islands however do not give rise to significant artifacts; they are generally extremely small. Only by extreme bad luck would the trajectory of interest, with an arbitrarily chosen stepsize, fall on one of these islands. We must however keep in mind this possible artifact. For example, each of these secondary unstable points is associated with its own chaotic zone. If we are calculating the Lyapunov exponent for our system, we should consider the possiblity that by chance we fell into one of these chaotic zones introduced by the mapping approximation.

We can say more clearly what it means for the mapping method to be non-linearly stable. If we could prove that the trajectory as computed by the map is an invariant curve or is bound in the phase space by invariant curves, then the trajectory will always remain just as far from the pendulum orbit as it was in the beginning. Of course, this is extremely difficult to do. The Kolmogorov-Arnold-Moser theorem proves that near the stable fixed point of the pendulum map that most of the invariant curves (with given frequency) of the pendulum are preserved in the mapping approxima-

tion. The exceptions are those which are too near the Poincare-Birkhoff islands, but these are of small measure. What happens at larger distances from the stable fixed point? Though it is not proven, numerical experiments suggest that the phase space is still dominated by invariant curves. One possibility, however, is that the invariant curves are actually cantori with small holes through which the mapping trajectory can ultimately leak out (Percival, 1979). Thus to prove the non-linear stability of the method for a particular trajectory for a particular stepsize, one must prove either that the trajectory is an invariant curve or that it is bounded by invariant curves rather than cantori. Practically though, the islands appear to behave as though most of the trajectories in them are invariant curves. An exception might be if the trajectory is very close to the edge of the island, for then the timescale to leak through the cantori may be comparable to the timescale of interest in the numerical experiment, but if the trajectory is close to the edge of the island then the stepsize has been chosen so that the mapping method is very close to the border of instability, as discussed above. This situation is easy to avoid, by choosing a better stepsize.

For systems with many degrees of freedom, the question of stability is much more complicated, principally because trajectories of the actual system can now be chaotic. We must consider what stability would mean for both quasiperiodic and chaotic trajectories. For quasiperiodic trajectories the situation is similar to the single degree-of-freedom case. If the mapping trajectory can be proven to lie on an invariant curve that remains close to the true invariant curve of the original system then the method can be declared to be stable. Of course, if we could prove these things we would not be using a mapping approximation to study the system. With a single degree of freedom, if the mapping trajectory is chaotic, but bound close to the true trajectory by invariant curves, then the mapping can be said to be stable. With many degrees of freedom, it is conjectured that all chaotic zones are connected (the "Arnold web") and that Arnold diffusion eventually carries the trajectory everywhere along it. Practically speaking though the Arnold diffusion is usually extremely slow, and the connectedness of the chaotic zones is often ignorable. Thus even if the mapping trajectory corresponding to a quasiperiodic trajectory of the modeled system

is chaotic, because for instance that it fell on one of the chaotic zones associated with one of the Poincare-Birkhoff unstable periodic orbits, we may still think of the mapping method as stable provided the computed trajectory stays near the actual trajectory for the duration of the numerical experiment. Of course the small positive Lyapunov exponent of the computed trajectory must be recognized as an artifact. It is also possible for the actual trajectory of the system to be chaotic. Stability of the mapping method in this case means that any quantity which characterizes the chaotic zone in which the trajectory moves is accurately reproduced by the mapping approximation. For example, it should be the case that the size and shape of the chaotic zone are well reproduced, as well as the Lyapunov exponent. It would not be expected that individual trajectories of the actual system and the map are the same. Rather, shadowing results suggest that computed trajectories approximate real trajectories of the original system, though they do not approximate trajectories which can be specified a priori (Grebogi, et al. 1990). We have argued that maps make sense from the point of view of averaging and this argument is born out by practical experience. The chaotic zones near the 3/1 resonance are well described by the mappings; the size and shape of the chaotic zones are reproduced, as well as the Lyapunov exponent (Wisdom, 1983). However, we do not know of rigorous results which show that chaotic trajectories of averaged systems shadow trajectories of unaveraged systems for long time. Nevertheless, it is plausible that the chaotic trajectories of the mapping approximations are satisfactory representations of the chaotic trajectories in the modeled system. Proving numerical reliability is much more difficult in this case. Thus, especially in more complicated problems, it is not likely that we can rigorously prove the stability of the method, especially since each trajectory and stepsize must be considered individually. We can however hope to determine the borders and mechanisms of the gross instabilities through the application of the more heuristic tools of non-linear dynamics such as the resonance overlap criterion for the onset of chaos.

We should not think of these possible artifacts as defects of the mapping method. The fact that the mapping method is itself a Hamiltonian system allows us to use the insight and methods of modern Hamiltonian dynamics to more clearly understand

the issues of nonlinear stability, as well as these possible artifacts. Surely similar artifacts would appear in any finite difference scheme for solving a system of ordinary differential equations; for example, stepsize resonances have also recently been found in the family of symmetric integrators (Quinlan and Tremaine, 1990, Quinlan and Toomre, 1992).

3.4 Stability of the N-Body Maps

The Hamiltonian nature of our mapping method allows us to study the non-linear stability of our mapping method using the usual tools of Hamiltonian dynamics. We propose that the principal instability of the mapping method is associated with an onset of chaos due to the overlap of resonances associated with the extra high-frequency terms introduced to generate the mapping. These resonances we call the stepsize dependent resonances. As usual in the analysis of the onset of chaos in a Hamiltonian system, we must first identify the principal resonances. We then calculate the location of each resonance and the width of the libration region associated with it. Chaos ensues if the principal resonance regions overlap. The resonance overlap criterion has been previously used in celestial mechanics to successfully predict the region of instability near the secondary in the restricted three-body body (Wisdom, 1980, Duncan, Quinn, and Tremaine 1989).

Resonances occur when linear combinations of the angular variables are slowly varying. When the Hamiltonian is written as a Poisson series in the angular variables, the most important resonances correspond to terms in the series with the largest coefficients. Writing the interaction Hamiltonian as a Poisson series in the Keplerian angle variables is the classical problem of the expansion of the disturbing function (e.g. Peirce, 1849, Plummer, 1960). An important property of this expansion is that the terms are proportional to various powers of the eccentricities and inclinations of the interacting planets; terms with arguments containing larger multiples of the longitudes of perihelion and longitudes of the ascending node are proportional to larger powers of the eccentricities and inclinations, respectively.

Expressing the mapping interaction Hamiltonian as a Poisson series is a straightforward extension of the usual expansion of the disturbing function. If we write the original interaction Hamiltonian as

$$H_{Interaction} = \sum_{\mathbf{i}} \beta_{\mathbf{i}} \cos(\mathbf{i} \cdot \theta), \tag{3.11}$$

where θ represents the full set of angle variables of the problem, and $\mathbf{i}$ is a vector of integers, $\mathbf{i} \cdot \theta$ then represents a particular angular argument of the Poisson series, and $\beta_{\mathbf{i}}$ represents its coefficient in the sum. In the mapping, the interaction Hamiltonian is multiplied by delta functions. The interaction Hamiltonian becomes

$$2\pi\delta_{2\pi}(\Omega t) H_{Interaction} = \sum_{\mathbf{i},l} \beta_{\mathbf{i}} \cos(\mathbf{i} \cdot \theta - l\Omega t). \tag{3.12}$$

That this equality holds may be seen by expanding the cosines of differences as a sum of a product of cosines and a product of sines. The sum over sines is zero because for every term with $l > 0$ there is an equal but opposite term with $l < 0$.

The most important resonances correspond to the terms in the Hamiltonian with the largest amplitudes. The largest terms in the expansion of the disturbing function are generally those with the smallest number of factors of the eccentricities and inclinations, since planetary eccentricities and inclinations are generally small. We consider only the lowest order terms. In the disturbing function itself there is only a single collection of terms which have no factors of eccentricity and inclination. The arguments of these terms are multiples of the difference between the mean longitudes of the pair of planets under consideration. These terms are not resonant except when the mean motions of the two planets are equal, but in this case the usual expansion of the disturbing function is not valid. Other terms in the expansion of the disturbing function are smaller because they are multiplied by various factors of the eccentricities and inclinations, but often are more important because in certain regions of the phase space their arguments can be slowly varying and there are large resonance effects. In the case of the mapping Hamiltonian, however, the terms in the disturbing function containing only pairwise differences of mean longitudes are now combined with multiples of the mapping frequency. The new combinations can be resonant. These are

the dominant stepsize dependent terms in the mapping interaction Hamiltonian; they control the basic stability of the mapping method.

We analyze the resonances resulting from the stepsize dependent terms as we would any other Hamiltonian resonance. We write the resonance Hamiltonian as the unperturbed Hamiltonian plus those interaction terms corresponding to the resonances of interest. We presume for the moment that the non-resonant terms can be pushed to higher order by some suitable canonical transformation. In this analysis, we also ignore any physical resonances that may also exist in the system. Thus, the Hamiltonian governing the dominant stepsize dependent resonances is given by

$$H_{Stepsize} = H_{Kepler} - \sum_{0<i<j<n} \frac{Gm_i m_j}{a_>} \sum_{k=0}^{\infty} \beta_k(\alpha) \sum_{l=-\infty}^{\infty} \cos[k(\lambda_i - \lambda_j) - l\Omega t], \quad (3.13)$$

where

$$\beta_k(\alpha) = \begin{cases} b^k_{1/2}(\alpha) & k > 1 \\ b^k_{1/2}(\alpha) - \alpha & k = 1 \\ \frac{1}{2} b^k_{1/2}(\alpha) & k = 0 \end{cases} \quad (3.14)$$

The sum over i and j is a sum over distinct pairs of planets, the number of planets being $n-1$. The masses of the planets are m_i. The semimajor axes are a_i, $\alpha = a_</a_>$, $a_>$ is the larger of a_i and a_j, and $a_<$ is the smaller of the two. The sum over k gives the set of terms in the disturbing function which are independent of the eccentricities and inclinations, the coefficients of which are given in terms of the usual Laplace coefficients $b^k_s(\alpha)$, and the arguments of which depend solely on the pairwise differences of the mean longitudes. There is a contribution from the indirect part of the disturbing function for $k = 1$. The sum over l comes from the Fourier representation of the Dirac delta functions.

Resonances occur when one of the angular arguments is nearly stationary. In considering which resonances are important we must keep in mind that the mapping frequency is larger than all the orbital frequencies. We found empirically that the mapping method performed well provided that 10 or more mapping steps were taken for each orbit period. Thus the mapping frequency Ω is larger than all of the mean motions (the orbital frequencies) by a considerable factor. Stepsize resonances occur

only if k is rather large. The coefficients are proportional to the Laplace coefficient $b^k_{1/2}(\alpha)$, which for small α is proportional to α^k. Terms with very high k are not as important as those with lower values of k. The most important stepsize resonances are those for which $l = \pm 1$. In any particular region of phase space, terms with larger $|l|$ have proportionately larger $|k|$.

In the derivation of the mapping method we used canonical Cartesian Jacobi coordinates and conjugate momenta (see WH91). Resonance analysis is more easily carried out in some form of canonical Keplerian elements. The most convenient set for the present purpose is the set of modified Delaunay elements which have as coordinates the mean longitude, the longitude of pericenter, and the longitude of the ascending node (see Plummer, 1960). Now our resonance Hamiltonian depends only on the mean longitudes, and the semimajor axes. The momentum conjugate to the mean longitude λ_i is $L_i = \sqrt{m'_i \mu_i a_i}$, where m'_i is the i^{th} Jacobi mass, which is nearly equal to m_i since the planetary masses are small, and $\mu_i = G m_i m_0$, where m_0 is the mass of the dominant central object. In terms of these canonical elements, the Keplerian Hamiltonian is

$$H_{Kepler} = -\sum_{i=1}^{n-1} \frac{m'_i \mu_i^2}{2L_i^2}. \tag{3.15}$$

In these variables, the Keplerian Hamiltonian is obviously integrable, since no angles appear.

Having identified the set of resonances of interest, we now analyse each in detail. For simplicity, we consider first only the terms for which $l = \pm 1$, the most important terms; the generalization to $|l| > 1$ will be immediate. Thus for each pair of planets we consider each term in the sum over k separately. We call this the k^{th} stepsize resonance. For definiteness, we shall assume $n_i > n_j$, i.e. that $a_i < a_j$. The Hamiltonian for the k^{th} stepsize resonance for planets i and j is then

$$H_{ijk} = -\frac{m'_i \mu_i^2}{2L_i^2} - \frac{m'_j \mu_j^2}{2L_j^2} - \frac{G m_i m_j}{a_j} \beta_k(\alpha) \cos[k(\lambda_i - \lambda_j) - \Omega t]. \tag{3.16}$$

This resonance Hamiltonian contains only a single linearly independent combination of the angles. As usual we make a canonical transformation to a resonance variable which is this sole combination of angles. The transformation is carried out with a

generating function of the form

$$F(\lambda_i, \lambda_j, \Lambda, \Sigma, t) = [k(\lambda_i - \lambda_j) - \Omega t]\Sigma + (c\lambda_i + d\lambda_j)\Lambda, \tag{3.17}$$

where the constants c and d are subject only to the constraint that $c\lambda_i + d\lambda_j$ be linearly independent from the combination $k(\lambda_i - \lambda_j)$. We arbitrarily choose $c = d = 1$. The new angle variables are the resonance variable $\sigma = \partial F/\partial\Sigma = k(\lambda_i - \lambda_j) - \Omega t$ and the fast variable $\lambda = \partial F/\partial\Lambda = \lambda_i + \lambda_j$. The relationships among the momenta are $L_i = \partial F/\partial\lambda_i = \Lambda + k\Sigma$ and $L_j = \partial F/\partial\lambda_j = \Lambda - k\Sigma$. The new Hamiltonian is

$$H'_{ijk} = H_{ijk} + \frac{\partial F}{\partial t} \tag{3.18}$$

$$= -\frac{m'_i\mu_i^2}{2(\Lambda + k\Sigma)^2} - \frac{m'_j\mu_j^2}{2(\Lambda - k\Sigma)^2} - \frac{Gm_im_j}{a_j}\beta_k(\alpha)\cos(\sigma) - \Omega\Sigma, \tag{3.19}$$

where a_j and α are assumed to be written in terms of the new momenta. This Hamiltonian now has only a single angle variable, σ, and is cyclic in λ. Thus the momentum conjugate to λ is an integral of the resonance Hamiltonian. The orbits of the conjugate pair (σ, Σ) can be determined simply by drawing contours of the resonance Hamiltonian H'_{ijk}, upon fixing the other constants.

Though the resonant motion is completely described by the contours of H'_{ijk}, it is instructive to study an approximation to it. We anticipate that the variations of the system away from exact resonance will be small. The resonance condition is that a difference of the mean motions of a pair of planets when multiplied by a relatively large integer is the mapping frequency. The fact that the coefficient β is small for large k suggests that only if the match of mean motions is rather good will there be any resonance effect. This will be confirmed by the following discussion. Thus, we assume the variations of Σ about some center Σ^* are small, and we expand the resonance Hamiltonian about this center. The σ dependent term is already small and to a good approximation it is well represented by its value at the resonance center. Define H_0 to be the σ independent terms

$$H_0 = -\frac{m'_i\mu_i^2}{2(\Lambda + k\Sigma)^2} - \frac{m'_j\mu_j^2}{2(\Lambda - k\Sigma)^2} - \Omega\Sigma. \tag{3.20}$$

Now expand H_0 about the resonance center

$$H_0 = H_0|_{\Sigma^*} + \left.\frac{\partial H_0}{\partial \Sigma}\right|_{\Sigma^*} (\Sigma - \Sigma^*) + \frac{1}{2}\left.\frac{\partial^2 H_0}{\partial \Sigma^2}\right|_{\Sigma^*} (\Sigma - \Sigma^*)^2 + \cdots. \tag{3.21}$$

The first term is a constant and can be ignored. The second term can be used to define the resonance center. At resonance the time derivative of σ is near zero. The time derivative of σ is given by Hamilton's equations as the derivative of the Hamiltonian with respect to Σ. Since the σ dependent term is small, the time derivative of σ is dominated by the derivative of H_0 with respect to Σ. Thus it is natural to define the resonance center Σ^* to be that point at which $\partial H_0/\partial \Sigma = 0$. Variables evaluated at the resonance center will be denoted by a superscript $*$. In more detail, the resonance center Σ^* is defined by

$$\left.\frac{\partial H_0}{\partial \Sigma}\right|_{\Sigma^*} = \frac{k m_i' \mu_i^2}{(\Lambda + k\Sigma^*)^3} - \frac{k m_j' \mu_j^2}{(\Lambda - k\Sigma^*)^3} - \Omega = 0. \tag{3.22}$$

Written in terms of the mean motions this resonance condition is just $k(n_i^* - n_j^*) - \Omega = 0$, as is easy to verify. As Λ is varied there is a continuous family of resonance centers. That there is a continuous family of resonance centers is easy to see: if for some n_i^* and n_j^* there is a resonance, then if both are increased by the same amount the resonance is maintained since the resonance condition depends only on their difference.

Returning now to the expansion of the resonance Hamiltonian about the resonance center, the linear term vanishes through the definition of the resonance center. We are left with the quadratic term in $\Delta\Sigma = \Sigma - \Sigma^*$ and the σ dependent term

$$H'_{ijk} = \frac{1}{2}\gamma(\Delta\Sigma)^2 + \beta\cos(\sigma), \tag{3.23}$$

where the second derivative of H_0 is

$$\gamma = \left.\frac{\partial^2 H_0}{\partial \Sigma^2}\right|_{\Sigma^*} \tag{3.24}$$

$$= -3k^2\left[\frac{m_i' \mu_i^2}{(\Lambda + k\Sigma^*)^4} + \frac{m_j' \mu_j^2}{(\Lambda - k\Sigma^*)^4}\right] \tag{3.25}$$

$$= -3k^2\left[\frac{1}{m_i' a_i^{*2}} + \frac{1}{m_j' a_j^{*2}}\right] \tag{3.26}$$

and β represents the coefficient of the σ dependent term

$$\beta = -\frac{G m_i m_j}{a_j^*}\beta_k(\alpha^*). \tag{3.27}$$

The width of the resonance is determined by tracing the separtrix. For $\beta < 0$ and $\gamma < 0$, the unstable equilibrium is at $\Delta\Sigma = 0$ and $\sigma = 0$, so the value of the Hamiltonian on the separatrix is β. The resonance is widest at $\sigma = \pi$ at which point $\Delta\Sigma = 2\sqrt{\beta/\gamma}$. This width corresponds to the maximum deviation from the resonance center for which libration is allowed. In terms of the modified Delaunay variables the half width is $|\Delta L_i| = |\Delta L_j| = k\Delta\Sigma = 2k\sqrt{\beta/\gamma}$. Note that because of the constraint that $\Lambda = (L_i + L_j)/2$ is constant the maximum of L_i corresponds to the minimum of L_j, and vice versa. For small resonance widths we can approximate the width in $\alpha = a_i/a_j$, by $\Delta\alpha/\alpha = 2(\Delta L_i/L_i - \Delta L_j/L_j) = 2(1 + L_i/L_j)\Delta L_i/L_i$.

We have confirmed the existence of these stepsize resonances by computing the evolution of trajectories of a two planet system. To be specific, we chose the masses and semimajor axes to correspond to those of Jupiter and Saturn. We set the initial eccentricities and inclinations to zero to minimize the effect of physical resonances in the system. We then carried out a large number of short integrations (only 1000 iterations) with a large number of stepsizes and monitored the variations in energy. The resulting diagram (see Fig. 3-1) displays a resonance structure which closely matches the predicted locations of the stepsize resonances. The relative energy variations are defined by $\Delta E = (E_{max} - E_{min})/(E_{max} + E_{min})$. Using this same system, we have also confirmed the detailed predictions of the analytical resonance analysis. We tuned the stepsize so that the system fell on the separatrix of the $k = 10$, $l = 1$ stepsize resonance. The resulting evolution is shown in Fig. 3-2. We plot the ratio of semimajor axes α versus the difference of mean longitudes of the two planets. We find a ten-lobed chain of islands, as expected. The width of each island, in α, agrees with the above resonance analysis to better than one part in a thousand. Variables referring to the inner planet with the mass of Jupiter are denoted by subscript 1; those referring to the body with the mass of Saturn by a subscript 2. The initial conditions are $\alpha = 0.54365$, $\lambda_1 - \lambda_2 = \pi/10$, to machine precision, and the stepsize is approximately given by $h/T_1 = 0.16701\ldots$, where T_1 is the orbital period of our Jupiter. The semimajor axis of our Jupiter is $a_1 = 5.2AU$, with $m_1 = m_{Sun}/1047.355$, and $m_2 = m_{Sun}/3501.0$. The length of the integration was about 20,000 years.

We now turn to the full ensemble of stepsize resonances. Fig. 3-3 displays the positions and widths of the stepsize resonances for $k = 2-30$, with $l = 1$. There are an infinite number of resonances in the lower part of the figure for larger k. The stepsize resonances accumulate both near zero stepsize, and for α near 1. As the stepsize goes to zero the linear density of resonances is proportional to $1/k^2$, but the widths of these resonances decreases exponentially as $\alpha^{k/2}$. They decrease in size much more rapidly than they accumulate. Thus this region does not give resonance overlap. As expected, small stepsizes are stable. This large number of very small stepsize resonances corresponds to a subset of the multitude of small Poincare-Birkhoff islands we saw in the standard-map analogy given above. The accumulation of resonances as α approaches 1 is indicative of a real integrator instability. The widths increase as the stepsize gets larger, so for any α there is a critical stepsize above which the mapping method is unstable. In Fig. 3-4, we show the same diagram when resonances with $l < 4$, and $k \leq 30$ are included. The higher order stepsize resonances fill in the gaps somewhat. Keep in mind that this figure has been computed for masses appropriate for the Jupiter-Saturn system; for other masses the diagram will be different.

We have tested the prediction of integrator instability in the region of resonance overlap, by computing the evolution of a large number of Jupiter-Saturn like systems with various stepsizes and initial semimajor axis ratios. The grid of initial stepsizes and semimajor axis ratios is quite fine, with 100 values in each parameter. Thus 10,000 experiments were carried out. Each integration was continued up to the point of dissolution of the system, or for 100,000 years, whichever came first. The cumulative time spanned by these test integrations is of order one billion years. This test of the map would have been unthinkable without the speed of the mapping method. The results are shown in Fig. 3-5. In this figure a point is plotted if during the integration one of the planets became hyperbolic. Lyapunov exponents were also computed; the regions of rapid divergence of trajectories were essentially identical to the regions of gross instability. We see that the agreement of the unstable regions with the regions of resonance overlap is quite good. The tongues of instability that reach down from the right part of the diagram correspond well with the tongues of resonance overlap.

Including higher order resonances and the chaotic width of the separatrices themselves would surely improve the agreement, but these are unnecessary refinements. The major instability is clear.

The experiment also reveals a second instability, unrelated to resonance overlap of the stepsize resonances, when the mapping stepsize is precisely half the orbital period of the innermost planet. Evidently, it is necessary for the stepsize to avoid low order commensurabilities so that the dynamics can accomplish a natural averaging over the orbit.

3.5 Conclusions

We have examined the stability of the new symplectic n-body maps from the point of view of non-linear dynamics. We have identified the resonances responsible for the principal artifacts. These are resonances between the stepsize and the difference of mean motions between pairs of planets. For large stepsizes resonant perturbations are evident in the variation of the energy of the system corresponding to these stepsize resonances. We have shown that the principal instability of the method can be predicted and corresponds to the overlap of the stepsize resonances. We note that the analysis suggests other artifacts will occcur. For example the overlap of a stepsize resonance with a resonance of the actual system may also give a region of chaotic behavior that is an artifact. We point out that the fact that the principal artifacts correspond to a particular set of stepsize resonances suggests that it may be possible to perturbatively remove the effect when the stepsize resonances are non-overlapping (see Tittemore and Wisdom, 1989).

3.6 References

Chirikov, B.V. (1979). Phys. Rep. **52**, 263.

Duncan, M., Quinn, T., and Tremaine, S.D. (1989). *Icarus* **82**, 402.

Grebogi, C., Hammel, S., Yorke, J., and Sauer, T. *Phys. Rev. Lett.* **65**, 1527.

Laskar, J. (1989). *Nature* **338**, 237.

Lichtenberg, A.J. and Lieberman, M.A. (1983). *Regular and Stochastic Motion*, (Springer-Verlag, New York).

Percival, I. (1979) in *Nonlinear Dynamics and the Beam-Beam Interaction*, M. Month and J.C. Herrera, Eds. (Amer. Inst. Phys., New York).

Peirce, B. (1849). *Astron. J.* **1**, 1.

Plummer, H.C. (1960). *An Introductory Treatise on Dynamical Astronomy*, (Dover, New York).

Quinlan, G.D. and Toomre, A. (1992). preprint.

Quinlan, G.D. and Tremaine, S.D. (1990). *Astron. J.* **100**, 1694.

Sanz-Serna, J.M. and Vadillo, F. (1986) in *Numerical Analysis*, Griffiths, D.F. and Watson, G.A., Eds. (Wiley, New York,)

Sussman, G.J. and Wisdom, J. (1992) submitted.

Tittemore, W. and Wisdom, J. (1987) *Icarus* **74**, 172.

Wisdom, J. (1980). *Astron. J.* **85**, 1122.

Wisdom, J. (1982). *Astron. J.* **87**, 577.

Wisdom, J. (1983). *Icarus* **56**, 51.

Wisdom, J. and Holman, M. (1991). *Astron. J.* **102**, 1528.

3.7 Figures

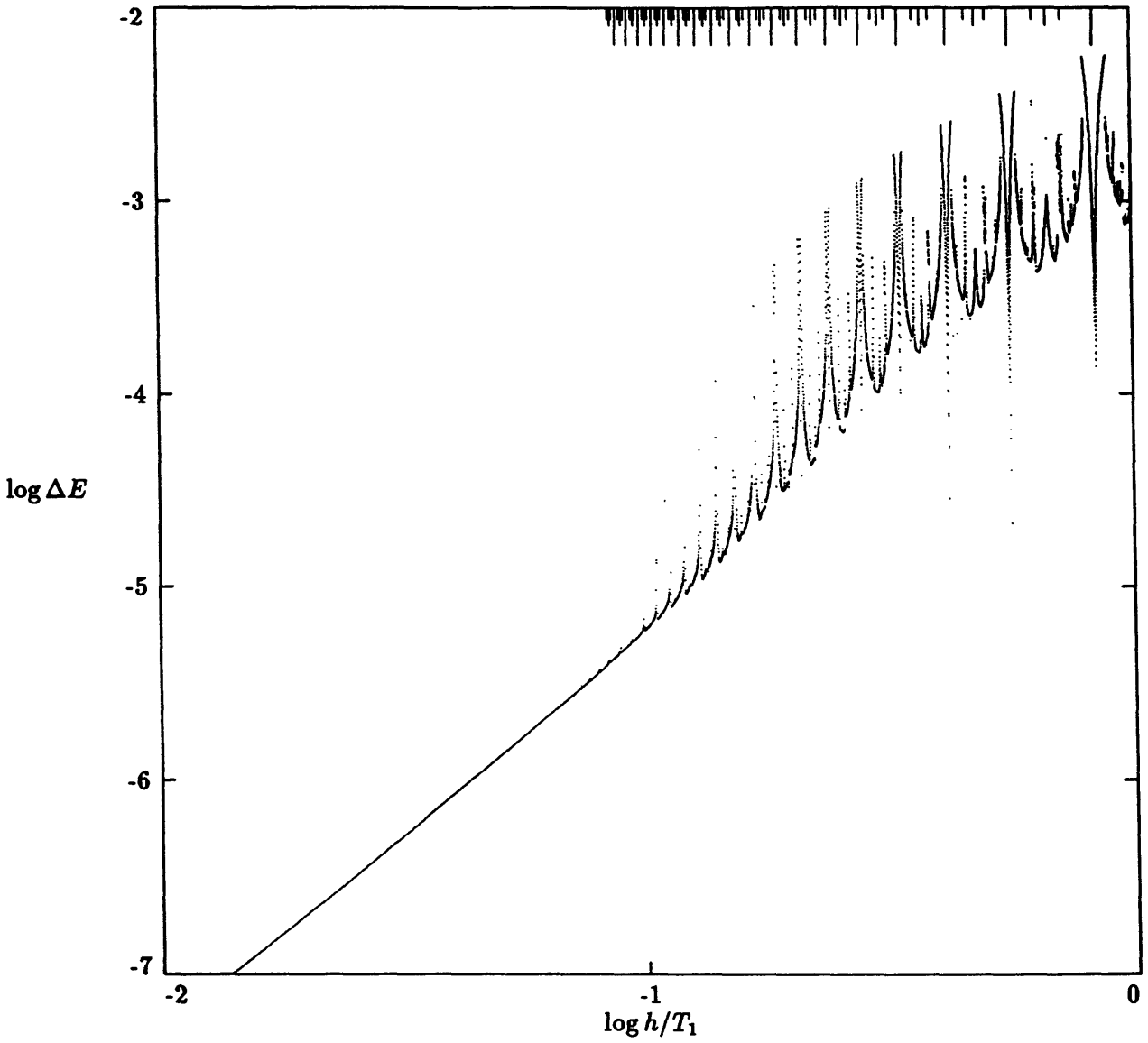


Figure 1: The mapping exhibits stepsize resonances as predicted by the theoretical analysis. The points present observed short term relative energy variations for a large number of stepsizes. The lines at the top mark the location of the stepsize resonances in the region where the effect on the energy is greatest. Shorter lines correspond to higher order stepsize resonances.

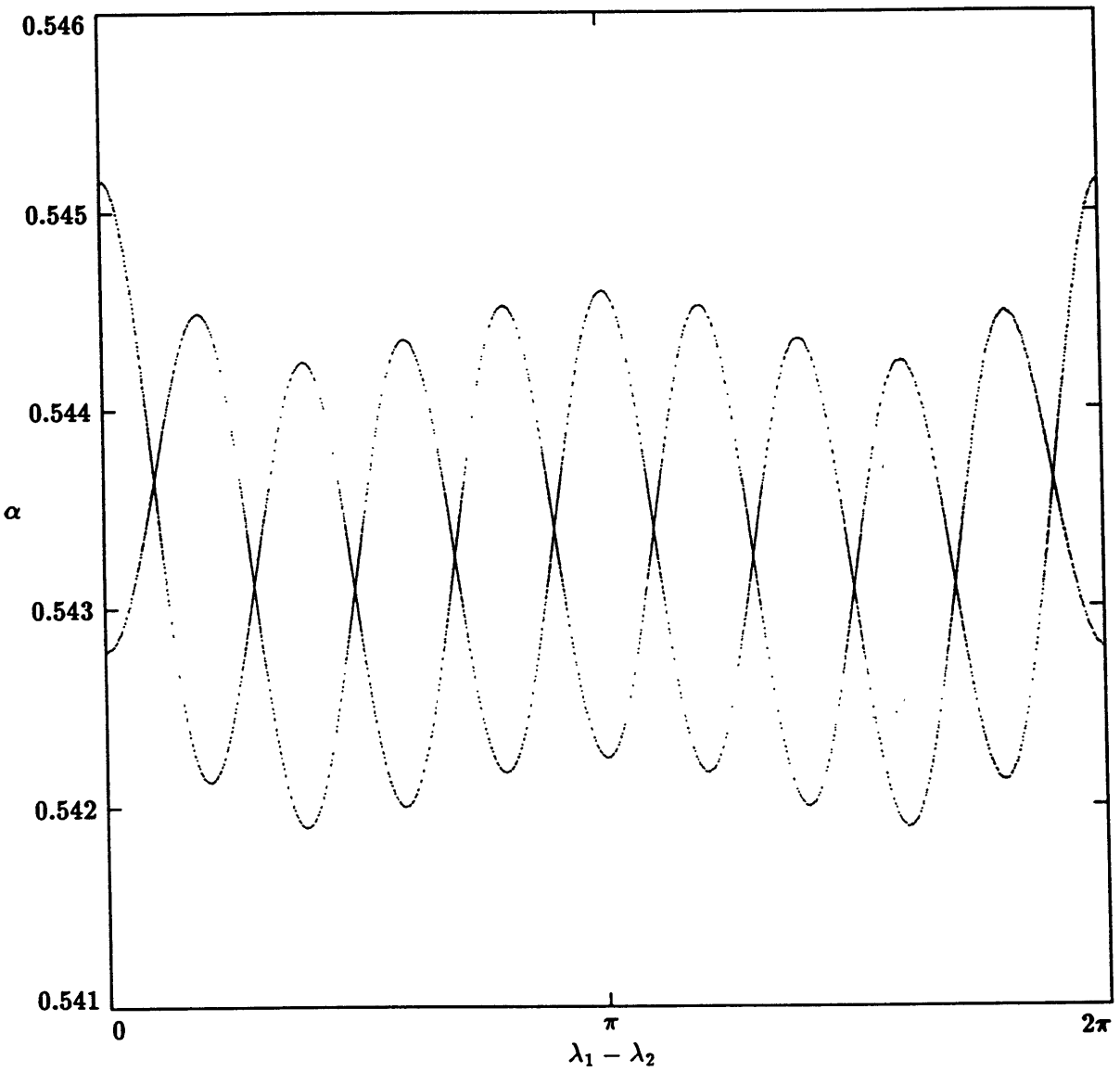


Figure 2: The stepsize resonances have the width predicted by the theoretical analysis. The evolution of a trajectory near the separatrix of the $k = 10$, $l = 1$ stepsize resonance is shown.

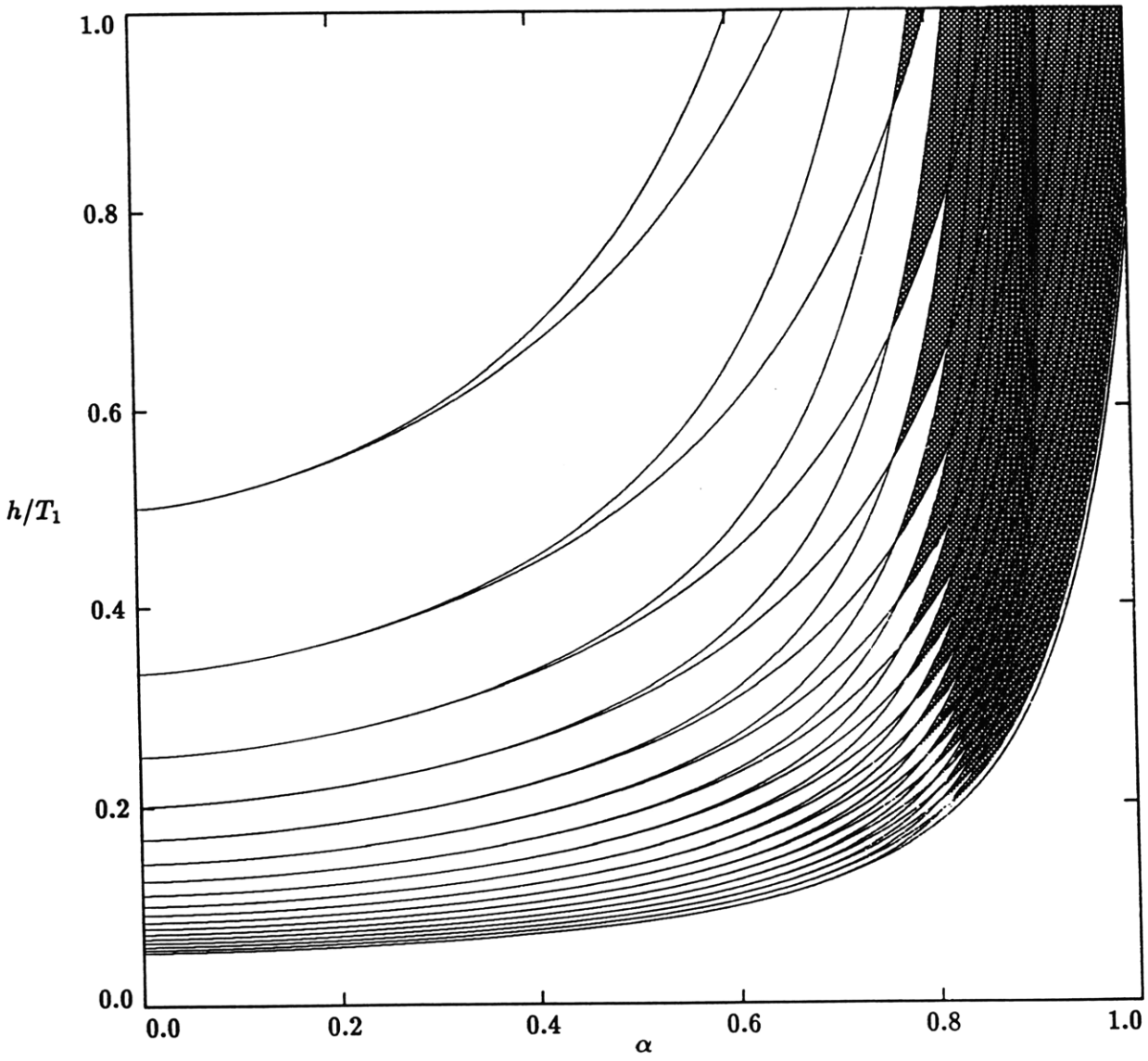


Figure 3: The positions and widths of the stepsize resonances with $k \leq 30$, $l = 1$. The infinity of resonances for larger k fall in the empty region in the lower right. For small α and h/T_1 the resonances get smaller faster than they accumulate. The region of resonance overlap is shaded. The overlap of resonances for α near 1 and moderate h is indicative of a real integrator instability.

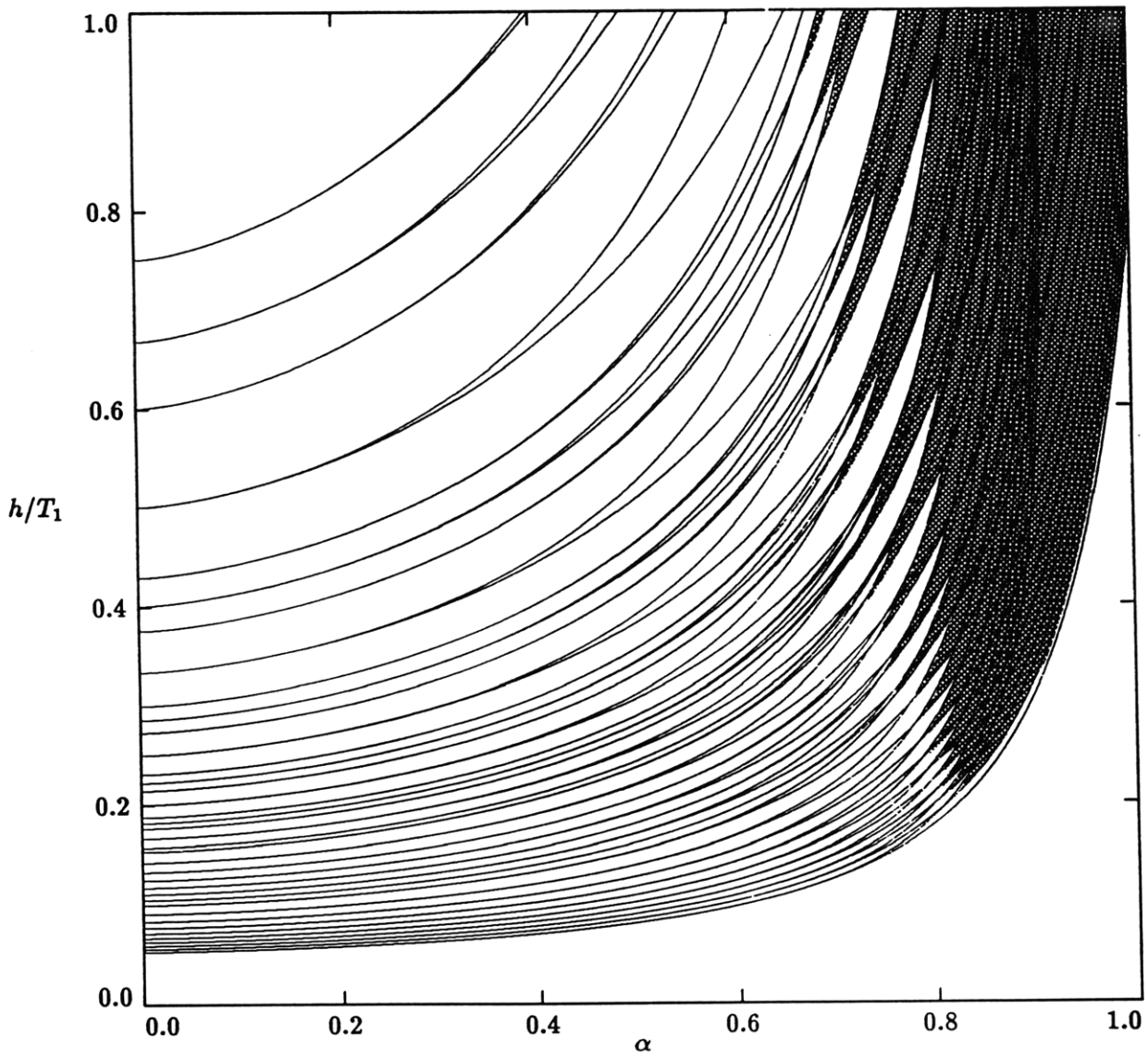


Figure 4: The same as Fig. 3, but including stepsize resonances with $l < 4$.

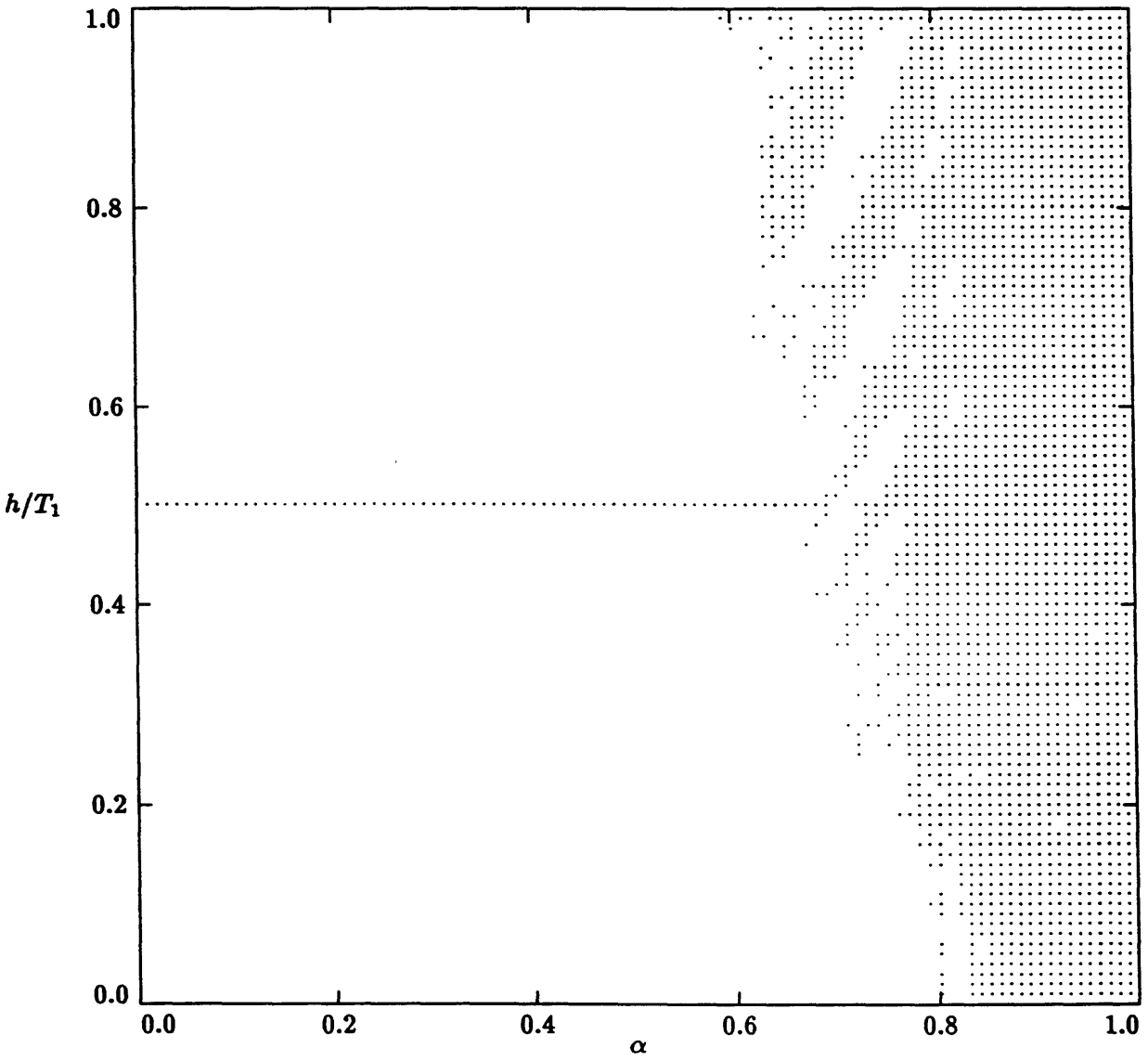


Figure 5: Unstable integrations are indicated by a dot. The agreement with the predictions of the resonance overlap of stepsize resonances is quite good. The overlap of first order resonances of the real system accounts for the physical instability at small h for $\alpha > 0.8$.

Chapter 4

Dynamical Stability in the Outer Solar System and the Delivery of Short Period Comets [3]

4.1 Introduction

Were there initially asteroids beyond Jupiter? Are there regions between the giant planets in which small bodies are stable against planetary encounters for the age of the solar system? Do Saturn, Uranus, and Neptune have Trojan-like asteroids? Are planetesimals in the hypothesized Kuiper belt stable against close encounters for the age of the solar system or is this region already depleted? On what timescale is material removed from different regions of the solar system? These questions motivate our survey of the long-term stability of small bodies in the outer solar system. In this study we extensively examine the stability of test particles in the regions between the outer planets and beyond Neptune.

In Section II, we review observational searches for slow-moving objects in the outer solar system. Section III reviews previous test particle surveys. Section IV describes the method of this survey. Sections V presents the details and results of our survey of the regions near the triangular Lagrange points of Jupiter, Saturn, Uranus, and Neptune. Section VI presents the results of our survey of the invariable plane. Section VII presents a summary.

[3] Adapted from Holman, M. and J. Wisdom (1993). Dynamical Stability in the Outer Solar System and the Deliver of Short Period Comets. *Astron. J.* **105**, 1987–1999.

4.2 Observational Searches

A number of observational surveys to detect slow-moving objects in the outer solar system have been conducted. Tombaugh (1961), Kowal (1989), Luu and Jewitt (1988), and Levison and Duncan (1990) have all conducted such investigations. In reviewing these investigations we consider in which regions each survey could have detected 624 Hektor, a bright Jupiter Trojan asteroid. At $H = 7.49$, 624 Hektor would be roughly $V = 17$, $V = 20$, and $V = 22$ if it were at the distance of Saturn, Uranus, or Neptune, respectively (Innanen and Mikkola, 1989).

Tombaugh examined the ecliptic region to a limiting magnitude of $m_{pg} = 17$ for slow-moving objects. Although he searched the Lagrange points of Saturn, his survey might have just missed the hypothetical bright Trojan at that distance. Other than the discovery of Pluto in 1930, no outer solar system bodies were found.

Kowal photographed about 6400 square degrees of ecliptic region, to a limit of $V = 20$, for slow-moving objects. Only one object was found beyond Saturn, 2060 Chiron. At V=17 magnitude and 17 AU from the Sun at the time of its discovery, Chiron was 3 magnitudes above Kowal's detection limit. Chiron is large (100 km) and cometary (Meech and Belton 1990). Its orbit is planet crossing, highly eccentric, and highly inclined. We have not yet determined if Kowal actually examined the triangular Lagrange points, however, if he did, it would have been possible to detect 624 Hektor at the distance of Saturn and possibly Uranus.

Luu and Jewitt (1988) surveyed 297 square degrees near the ecliptic to a limiting magnitude of $V \approx 20$ using Schmidt plates and 0.34 square degrees to a limit of $R \approx 24$ with a CCD camera. Like Kowal's, their plate survey could detect 624 Hektor at distances of Saturn and possibly Uranus. With the CCD survey they could detect the test object even at Neptune distances. At the times of Luu and Jewitt's observations, from February to June, 1987, Saturn, Uranus, and Neptune were quite close together. In fact, the L_5 points of these three planets were near opposition. Consequently, the centers of about two thirds of the fields observed by Luu and Jewitt are within 30°, along the ecliptic, of the L_5 points of Saturn, Uranus, or Neptune. This amounts to

observations of about 200 square degrees near the L_5 points observed to $V \approx 20$. A few of the fields are quite close to the L_5 points. Considering the range of libration observed for stable test particles in our integrations, their choice of fields is appropriate for detecting objects at the Lagrange points. However, they found no slow-moving objects in this search.

Finally, Levison and Duncan (1990) examined 4.9 square degrees of the ecliptic, explicitly limiting their survey to a search for slow-moving objects in the range 25-60 AU. Their detection limit was $V \approx 22.5$. 624 Hektor at the distance of Neptune would be very near the detection limit. Detecting nothing, they report a 99% confidence level result that fewer than one object per square degree brighter than $V \approx 22.5$ lies between 25 AU and 60 AU. (Does this hold up?) Of the 26 fields examined 8 lie within 30° of the L_4 or L_5 points Neptune. This amounts to observations of about 1.5 square degrees near Neptune's triangular Lagrange points. Several more fields are within 45°, still within range of the widest librators. Even so, this is not a great deal of area considering the range of libration.

Since the results of these surveys were first reported several Kuiper belt candidate objects have been discovered, beginning with the discovery of 1992 QB1 by Luu and Jewitt (1992). To date, a total of eight have been listed. The discovery circumstances, prelimary elements, and magnitude estimates are listed in Table 4.1. Only one of these objects (1993 SC) has V brighter than the $V \approx 22.5$ detection limit of Levison and Duncan. Considering that only one object that bright has been reported and that more than one square degree has been surveyed, the reported limit of Levison and Duncan (1990) on the number of bright objects between 25 AU and 60 AU seems to hold.

In summary, the surveys of Tombaugh, Kowal, and Luu and Jewitt could all have detected bright Saturn Trojans. Some of the observations, including those of Levison and Duncan, could have detected bright objects at the distances as far as Neptune. To date, eight Kuiper belt candidate objects have been discovered. The elements for these objects are uncertain. Most of them have only been observed at one opposition.

4.3 Previous Test Particle Surveys

There have been test particle surveys of both the Lagrange points of the outer planets and the regions between and beyond the outer planets. The Trojan asteroids occupy the regions near Jupiter's Lagrange points. Similar Trojan-like configurations have been seen in satellite systems, and the asteroid 1990 MB appears to be a Mars Trojan (Levy and Holt 1990, Bowell 1990, Kinoshita 1990). Yet, are the L_4 and L_5 points of other planets stable? Can material placed around these points remain there for the age of the solar system, or will the perturbations of the other planets induce its removal? Analytic methods are inadequate to answer these questions. We must rely on numerical exploration to investigate the long-term stability in realistic models. Levison *et al.* (1991) studied the stability field of the Jupiter Trojans. They integrated 110 test particles in the field of the Sun and four Jovian planets with a fourth-order symplectic integrator. They explored a two-dimensional grid of proper eccentricity and libration amplitude (see Shoemaker *et al.* 1989 and Erdi 1978, 1979). Their integrations extend to 150,000 Jupiter periods or about 1.8 million years. Zhang and Innanen (1988a,b,c) and Innanen and Mikkola (1989) investigated the stability of the triangular Lagrange points of all the Jovian planets. They studied the evolution of a few test particles for 10 million years, subject to the perturbations of the four Jovian planets. They found that test particles placed at the triangular Lagrange points of Jupiter, Uranus, and Neptune survived without close encounter for 10 million years. However, test particles placed at Saturn's L_4 and L_5 points approach Saturn on short timescales; whereas those initially placed a small distance away from the Lagrange points librate without close approach. Recently, Mikkola and Innanen (1992) provide a detailed description of the evolution of a number of orbits over 20 million years. We build upon the work of Innanen and Mikkola (1989, 1992) in the context of our model.

There have been a number of previous surveys of test particle stability in the regions between the outer planets, most of which focus on the region between Jupiter and Saturn (Lecar and Franklin 1973; Franklin, Lecar, and Soper 1989; Soper,

Franklin, and Lecar 1990; Weibel, Kaula, and Newman, 1990; Duncan, Quinn, and Tremaine 1989, hereafter **DQT89**; Gladman and Duncan 1990, hereafter **GD90**). Various models have been used to study this region. These models range in complexity from planar models that include Jupiter and Saturn following circular orbits to three-dimensional models that permit Jupiter and Saturn to fully interact. As the models have become more realistic a general consensus on this region has emerged. It is observed that most test particles develop planet crossing orbits or suffer a close encounter with Jupiter or Saturn within 10^4 to 10^5 years; a few test particles survive longer, 10^6 years. Our study is the first to investigate the Jupiter-Saturn region including all four giant planets in a self-consistent n-body integration.

Fewer studies have examined the test particle stability in the regions between the other outer planets. **DQT89** studied the stability of test particles in the region 0.6-34 AU using a simplified two-planet mapping approach. With this method they examined the regions between each pair of adjacent planets, including only those two planets as perturbers and approximating the perturbations on the test particles as impulses at each conjunction. Their integrations extended to 4.5 billion years. Aside from including only the two adjacent planets as perturbers other approximations were: (1) the planets and test particles were coplanar; (2) the planets were restricted to fixed circular orbits; and, (3) the eccentricities of the test particles were assumed to be small. In the model of **DQT89** many nearly circular orbits in the Saturn-Uranus and Uranus-Neptune regions survive for the age of the solar system. **GD90** studied the invariable plane from 3-40 AU with direct numerical integration of the three-dimensional n-body equations of motion, with a fourth-order symplectic integrator (Candy and Rozmus 1990; Forest and Ruth 1990). For test particles beyond Saturn they included all four giant planets as perturbers. **GD90** integrated roughly a thousand test particles for periods up to 22.5 million years, removing any test particle which encountered a planet or left the system. **GD90** reach a very different conclusion from **DQT89**. **GD90** observe that almost all of the test particles on orbits between the planets are unstable against close encounters on a timescale of about 10 million years. Thus **GD90** found that **DQT89** overestimated the planet-

crossing times by a substantial factor. Our survey extends and refines the study of **GD90**.

The test particle stability of the region beyond Neptune has received even less attention. Duncan, Quinn, and Tremaine (1988, 1990) find that the distribution of the orbits of short period comets is more consistent with an origin in the hypothesized Kuiper belt of comets in the region beyond Neptune than with an origin in the isotropic Oort cloud. However, the model from which this result emerged has a questionable feature: in order make their study computationally feasible the masses of the planets were enhanced by a considerable factor. This may be adequate for demonstrating that short period comets more likely come from a low inclination source region beyond Neptune, but is surely inadequate to evaluate the dynamics prior to planetary encounters. **DQT89** and **GD90** integrate a small number of test particles initially beyond Neptune with models that do not rely on enhancing the planetary masses. However, both studies report that the test particles, except for those quite close to Neptune, retain nearly circular orbits for the duration of their integrations. Torbett and Smoluchowski (1990, hereafter **TS90**) more extensively examined the evolution of test particles beyond Neptune using conventional numerical integration. However, in their model the outer planets moved on fixed elliptical orbits. They followed the evolution of about 200 test particles, 40 of which were initially placed on nearly circular orbits, for 10 million years. **TS90** discovered a chaotic zone beyond Neptune which corresponds roughly to orbits with perihelia between 30 and 45 AU. They observed Neptune crossing for only the test particles with initial perihelia quite close to the orbit of Neptune, within 2 AU. They noted that several test particles seemed to random walk through $a - e$ space approximately along lines of constant perihelion. They suggest that the chaotic zone is connected and that orbits diffuse throughout it roughly maintaining constant perihelia. They conjecture that the larger semimajor axis portion of this chaotic zone provides a storage place for short period comets which were initially formed at low eccentricity. The diffusion of comets to Neptune crossing orbits implies an exponential decay in the number of comets stored in the chaotic zone. Clearly, the dynamics of the Kuiper belt deserves much more

extensive exploration in more realistic models.

4.4 Method

Our approach is simple and direct. We integrate the motion of test particles in the field of the Sun and massive outer planets, Jupiter through Neptune, with the symplectic mapping method of Wisdom and Holman (1991, hereafter **WH91**). The Sun, planets, and test particles interact in the full 3-dimensional n-body sense. The test particles have infinitesimal mass; they are perturbed by the massive planets, but do not perturb them in return. The initial positions and velocities of the planets and Sun are taken from Cohen, Hubbard, and Oesterwinter (1973, hereafter **CHO73**). The accuracy and stability of the symplectic mapping method are analyzed and discussed in Wisdom and Holman (1992.)

During the integration the test particles are examined at each time step for close encounters with planets; those which enter the sphere of influence of a planet are terminated. The sphere of influence or activity sphere is a measure of the distance from a planet within which it becomes reasonable to consider a test particle in orbit around the planet perturbed by the Sun, rather than in orbit around the Sun perturbed by the planet (see Danby 1988). The radius of the sphere of influence is

$$r_s = a\mu^{2/5}, \tag{4.1}$$

where a is the initial semimajor axis of the planet and μ is the ratio of the mass of the planet and that of the Sun. This criterion is slightly different from that used by **GD90**:

$$r_s = a(2\mu)^{2/5} \approx 1.32a\mu^{2/5} \tag{4.2}$$

We do not believe the exact size of the sphere is important to the qualitative results. Although a planetesimal could survive a close approach without cataclysm the orbital elements of the particle would be radically altered. In addition to close encounters, we examine the test particles for parabolic or hyperbolic orbits. During our integrations, no non-elliptic orbits were detected before a close encounter.

4.5 Survey of the Lagrange Points of the Outer Planets

In this segment of our survey we study the evolution, for intervals up to 20 million years, of about 4000 test particles distributed near the Lagrange points of the outer planets. The test particles are given the same eccentricity, inclination, longitude of ascending node, and mean anomaly as one of the Jovian planets. The argument of pericenter is offset from that of the planet by a wide range of angles. The initial semimajor axes of the test-particles are equal to that of the planet multiplied by a "semimajor axis factor" ranging from 0.96 to 1.04. In the initial survey the argument of pericenter was varied from 0 to 360 degrees with a step of 5 degrees; the semimajor axis factor was varied from 0.96 to 1.04 at steps of 0.01. Additional initial conditions were used to trace detail in the most interesting regions. This choice of initial conditions places the test particles initially in the plane of the corresponding planet. The idea of modifying the initial semimajor axis is due to Innanen and Mikkola (1989) who first used this technique to investigate the stability of test particles at Saturn's triangular Lagrange points. In the case of the particles with the same initial semimajor axis as the planet (semimajor axis factor = 1.0) and the argument of pericenter offset by 60° or −60° the test particle would maintain this configuration if the perturbations of the planets other than the one in question were neglected. Those which enter the sphere of influence of a planet or develop non-elliptic orbits are removed. A time step of 1.0 year is used. We record the range of the difference in mean longitude of each particle from its corresponding planet, as well as the range of semimajor axis, eccentricity and inclination. These statistics are updated after each ten time-steps.

To test that the time-step and interval between updating the range of elements were not too short we repeated one of the runs of Jupiter Trojans with a time-step of 0.5 years and five time-steps per statistics update. The results are qualitatively the same as the run with a 1.0 year time-step and 10 integration steps before updating the statistics.

In Figure 4-1 we plot a point for each test particle that survived the full integration,

20 million years. We plot the offset of the argument of perihelion from the planet as the initial longitude versus the semimajor axis factor. A stable region surrounds the triangular Lagrange points of each of the outer planets. Innanen and Mikkola (1989) earlier found that test particles near the triangular Lagrange points of Saturn, Uranus, and Neptune could endure integrations of up to 10 million years, but now we have a two-dimensional view of initial conditions for test particles that endure integrations of 20 million years without close encounters.

A collection of stable initial conditions surrounds each of the triangular Lagrange points, but this does not imply that the test particles are confined to this region. In fact, over the course of the integration even a test particle initially at the L_4 and L_5 points can explore a large range of longitude with respect to its corresponding planet, 35° for Neptune's L_4 point as an example. Those initially placed further from the Lagrange points explore a larger longitudinal range, in some cases 100°!

Figure 4-1 immediately raises some questions. What governs the profile of the stable regions? Why do Saturn's L_4 and L_5 stable regions have holes in the center where the others do not? Innanen and Mikkola (1989) suggest the near 5:2 resonance between Jupiter and Saturn as a possible cause for the instability near Saturn's triangular Lagrange points. What is the cause of the apparent asymmetry for the Neptune L_4 and L_5? Although we do not expect the stable regions to be precisely symmetric due to asymmetric planetary phases, the asymmetry is pronounced in the case of Neptune. Can we estimate the phase volume of the stable region in order to predict the likelihood of observing material at these points? Are these regions just seemingly stable and will disappear with further integration? Future effort is clearly needed, but, for now, it has been established that for a non-negligible range of initial conditions test particles near the L_4 and L_5 points of Saturn, Uranus, and Neptune, as well as Jupiter, can endure without close encounter for up to 20 million years.

4.6 Survey of the Invariable Plane

In this segment of our survey we place 3000 test particles on heliocentric circular orbits in the **CHO73** invariable plane. The initial longitude of each test particle is assigned one of six values: 0, $3\pi/10$, $7\pi/10$, $11\pi/10$, $15\pi/10$, and $19\pi/10$ radians, measured from the x-axis of the **CHO73** coordinate system. Along each of the six longitudes, we uniformly distribute 500 test particles in the range 5-50 AU. Using the symplectic mapping method described above, each test particle is evolved in the field of the Sun and Jovian planets. Interior to Neptune the integrations have been extended to 4.5 billion years, and exterior to Neptune the integrations have been advanced to 1 billion years. The integration time-step is 1.0 year. During the integration, values of the minimum and maximum semimajor axis, eccentricity, and inclination explored by each particle are updated after every one hundred mapping steps.

Figure 4-2 and Table 4.2 present results from our invariable plane integrations. In Figure 4-2 the time survived by each test particle is plotted as a function of initial semimajor axis for the full range of initial semimajor axes explored. As mentioned above each bin in semimajor axis contains 6 test particles started at different longitudes in the invariable plane. The vertical bars mark the minimum of the six termination times. The survival times of the remaining 5 test particles are marked by small dots. The points along the top represent test particles that survived the full integration.

There are several features in the Figure 4-2 to point out. The spikes at 5.2, 9.5, 19.2, and 30.1 AU, at the semimajor axes of the planets, correspond to test particles librating about the triangular Lagrange points. Surrounding each of these spikes is a range of semimajor axes in which test particles quickly encounter one of the planets. The width of these regions corresponds well with the range of semimajor axis near a planet within which a test particle is predicted to undergo large-scale chaotic motion due to the overlap of first-order mean motion resonances (Wisdom 1980, **DQT89**). The half-width is approximately

$$\Delta a \approx 1.5 a \mu^{2/7}, \tag{4.3}$$

where a is the semimajor axis of the planet and μ is the ratio of the masses of the planet and the Sun. **GD90** derive an expression for the range of semimajor axis in the restricted three-body problem for which the Jacobi constant permits initially circular orbits to become crossing orbits. Their expression for the half-width is

$$\Delta a \approx 2.1 a \mu^{1/3}. \tag{4.4}$$

Both expressions predict the region of rapid removals fairly well, though the estimate based on the chaotic zone is more generally valid for two reasons. First, the Jacobi constant alone does not determine whether particular orbits are quasiperiodic or chaotic. The restricted three-body problem has the usual divided phase space with a mixture of chaotic and quasiperiodic trajectories for each value of the Jacobi constant (Hénon, 1966). Orbits do not generally explore the full range of the phase space permitted by the Jacobi constant. In fact, the validity of the crossing zone calculation rests on the fact that the chaotic zone is larger than the crossing zone, implying that the orbits considered in the crossing zone calculation are chaotic and free to explore the phase space which happens to include the crossing orbits. Second, when the perturbations of all the planets are considered the Jacobi constant does not exist, and the calculation has no rigorous generalization. The only generalization is by analogy. On the other hand, the resonance overlap estimate of the extent of the chaotic zone is still valid in the more general case. Furthermore, it can presumably be made more and more accurate by considering higher order resonances with the perturbing planets. **GD90** choose not to examine regions within the crossing zones of the planets. Here, we have placed test particles from 5-50 AU, leaving no gaps. This is not a costly decision because the test particles in these zones are rapidly removed from the integration. For a small price we get a complete picture.

As can be seen in Figure 4-2 most of the test particles initially between Jupiter and Saturn have had close encounters within 10^4 to 10^5 years. By 10^6 years all the test particles in the Jupiter-Saturn region have been eliminated; most of the test particles in the region encounter Jupiter or Saturn; a few encounter Uranus and Neptune. This is the first time that the Jupiter-Saturn region has been examined with a n-

body calculation that includes all of the Jovian planets as perturbers. Despite the addition of Uranus and Neptune as perturbers, our results qualitatively agree with a number of other studies that employ different models and methods to investigate test particle stability in the Jupiter-Saturn region.

As an aside, **GD90** observe that test particles with moderate initial inclinations in the Jupiter-Saturn region begin to be removed by close encounters later than those initially in the plane, but are then removed by close encounters more rapidly. The end result is that by 10^5 years the number of remaining test particles is roughly the same regardless of initial inclination. This observation indicates that restricting the initial conditions of the test particles to the invariable plane should not unreasonably skew the overall results, although we would have to examine a larger portion of the phase space to verify this assumption.

After 4.5 billion years no test particles initially in the region between Saturn and Uranus remain; only 1 test particle between Uranus and Neptune survived the full integration. From the figure it can be seen that most test particles in the Saturn-Uranus and Uranus-Neptune regions are eliminated within 10^7 years. Except for a few test particles librating about Neptune's triangular Lagrange points, the region surrounding Neptune is rapidly depleted, roughly in accord with the extent of the chaotic zone described above. In the region further beyond Neptune, numerous particles have had close encounters within 1 billion years. Thus we are witnessing the depletion of a much larger region than has been seen before in shorter integrations.

TS90 observed Neptune crossing for particles with initial perihelia within 2 AU of Neptune in 10 million years. **GD90** integrated 20 test particles uniformly distributed in 32.8-40 AU for 22.5 million years. Three test particles between 32.8 and 33.7 AU encountered Neptune, but the other test particles, with semimajor axes between 33.7 and 40 AU, retained nearly circular orbits during their 22.5 million year integration. For comparison, we observe that 42 of 60 test particles in the range 32.8-33.7 AU had close encounters within 22.5 million years. Also, we observed that 23 of 420 test particles with initial semimajor axes between 33.7 and 40 AU had close encounters within 22.5 million years. If we had sampled at the rate of **GD90** we would expect

to have observed 2-3 close encounters in the range 32.8-33.7 AU and fewer than one close encounter beyond 33.7 AU. Furthermore, we observed no close encounters of test particles beyond 39 AU before 22.5 million years. Thus, our results are consistent with those of **GD90** and **TS90**; the depletion of the regions beyond 32 AU which we observe begins later.

Note that the regions in which both **GD90** and **TS90** find Neptune crossing are similar and correspond roughly to the chaotic zone surrounding Neptune predicted by the overlap of first-order mean motion resonances (Eq. 4.3). Beyond this region, those two studies did not observe test particles on initially near circular orbits encountering Neptune. Our study shows that integrations of only 10-20 million years are inadequate to begin to see the extensive depletion of the region beyond 33 AU. **GD90** suggest that there may be a slow outward erosion of the Kuiper belt. We observe that the inner edge of this disk is not simply eroded outward, rather the disk is undergoing a more extensive irregular depletion.

In connection with the removal of test particles in the semimajor axis range 40-42 AU, Figures 4-3 and 4-4 reveals a bump in maximum eccentricity and inclination attained by test particles. Without examining the dynamics more closely we note the 3:2 mean motion resonance with Neptune near 40 AU and secular resonances near 41 AU (Heppenheimer 1979, see also Knezevic, et al. 1991). We also note the 2:1 mean motion resonance with Neptune near 48 AU.

We observe the phenomenon reported by **TS90** that in many cases orbits appear to random walk in the space of semimajor axis and eccentricity roughly along lines of constant perihelion. However, we do not find that motion in the more extensive chaotic zone reported by **TS90** has this character. Rather, we find that particles typically retain a semimajor axis near the initial semimajor axis while the eccentricities grow irregularly. Only when the perihelia get in the vicinity of Neptune do the particles begin to random walk along lines of constant perihelia. Typically they then relatively quickly have close encounters with Neptune. More rarely particles then begin again to random walk along another path of near constant semimajor axis with varying eccentricity. There are two classes of pathways: those roughly preserv-

ing perihelion distance and those roughly preserving semimajor axis. The constant perihelion paths most often lead to close encounter, but occasionally also serve to connect paths which preserve semimajor axis. The process is quite reminiscent of Arnold diffusion, in which particles successively travel along chaotic zones associated with different resonances. The typical evolution is illustrated in Figure 4-5. In this figure lines of constant semimajor axis are diagonal; lines of constant perihelion are vertical. The diagonal portion of the trajectory is quite narrow. Figure 4-6 shows the eccentricity versus time for this trajectory. Thus our simulations suggest a different storage mechanism for the short period comets than suggested by **TS90**. As of 1 billion years, the particles having late encounters with Neptune come from the region beyond 32 AU and were stored in one of the semimajor axis preserving chaotic zones. We do not observe long storage in the constant perihelion chaotic zone as suggested by **TS90**. Thus the dynamics of the delivery of short period comets appears to be more analogous to the dynamics of delivery of meteorites from mean motion resonances and secular resonances in the asteroid belt (Wetherill, 1968; Wisdom, 1985; Froeschlé and Scholl, 1986; Wetherill, 1987). The eccentricity grows irregularly while the semimajor axis remains relatively unchanged.

As an aside, note that the opposite of the evolution seen in Figure 4-5 is possible. A small body in the vicinity of Neptune, possibly an escaped satellite, could evolve along a path with roughly constant perihelion, and subsequently transfer to one of the paths of constant semimajor axis, ultimately evolving to an orbit of low eccentricity with large semimajor axis. The scenario is provocative, although not necessarily of practical importance. Mikkola and Innanen (1992) recently noted that for some initial conditions test particles placed near Neptune's triangular Lagrange points temporarily develop orbits similar to that of Pluto before being ejected.

It is particularly noteworthy that the survival times for particles between adjacent planets can vary by more than two orders of magnitude. The profile is quite jagged. The profile of survival times beyond Neptune is obviously clipped by the limited time of our integrations; the profile suggests that we are just beginning to see the depletion of this region, and that the spread of survival times will be equally large. Thus it is

likely that particles on initially circular orbits encounter Neptune over a wide range of times from about 10 million years to, say, 10 billion years. The range of survival times probably encompasses the age of the solar system.

The plot of the number of particles remaining as a function of time is remarkable (Fig. 4-7). The population does not decay exponentially as might have been expected. Nor does it decay as a power law. The population is best described as decaying logarithmically! Thus equal numbers of particles encounter Neptune for the first time in equal intervals of the logarithm of time. There was no reason to expect any particular decay law; each particle has its own individual deterministic dynamics. The logarithmic decay is just an average description of the dynamics of a large number of particles. It is interesting to note than a logarithmic decay implies a maximum lifetime; there is a time beyond which all particles which will have encounters have had them. The logarithmic decay is also recognizable in Fig. 4-2; there are apparently about equal numbers of dots in equal intervals of the logarithm of the survival time. Otherwise stated, the flux of new Neptune encounters is decaying as $1/t$, where t is the time since formation. If this trend continues we can speculate that, for instance, the flux of short period comets 3 billion years ago was only about 4 times the flux today.

We can continue our speculation and estimate how many comets would have had to have been initially in the region between 30 and 50 AU to account for the observed flux of new short period comets. Presuming a flux of new short period comets of about 0.01 per year (Fernandez 1985; and Duncan, Quinn, Tremaine 1988); and presuming that 0.17 of comets which encounter Neptune become visible (Duncan, Quinn, Tremaine 1988), we find that there were initially about 4×10^9 comets in this region and that roughly half of these comets must still be there. Assuming an average mass per comet of $10^{14.5}$ kg (Duncan, Quinn, Tremaine 1988) the total mass in the belt from 30-50 AU is then about 0.1 Earth masses, which is within the upper limit of about 0.2 Earth masses placed on the mass of the Kuiper belt from modeling perturbations to the orbit of Halley's comet (Hamid *et al.* 1968, Hogg *et al.* 1991). Our estimate is surely quite crude; for instance the extrapolated flux could easily be

off by a factor of several, and our survey only considers initially circular orbits in the invariable plane. The estimate can be easily refined with longer integrations.

Lecar *et al.* (1992) found an interesting correlation between the maximum Lyapunov exponent and the planet crossing time for test particles in the outer asteroid belt and in the region between Jupiter and Saturn. They report a power law relation between the planet crossing time and the Lyapunov time. In their experiments, the best fit exponents relating the crossing time to the Lyapunov time are roughly 1.8. Figure 4-8 shows the correlation of the Lyapunov time (the inverse of the maximum Lyapunov exponent) and the time of close encounter for the test particles is our study. We find that the time of close encounter is more nearly directly proportional to the Lyapunov time, perhaps with an exponent as large as 1.4. A strong correlation could be used to estimate the times of close encounter for objects remaining in the region beyond Neptune based on the measured Lyapunov times, leading to another estimate of the flux from the region. However, the termination times range over two orders of magnitude for any given value of the Lyapunov time, limiting predictions of close encounter times to a rough range of values. A more direct approach is to improve our estimate of the flux is to extend the integrations to 4.6 billion years.

4.7 Summary

On timescales of twenty million years we find no evidence that Saturn, Uranus, and Neptune cannot retain Trojan-like asteroids. Test particles in Saturn's orbit plane and placed on orbits near Saturn's L_4 and L_5 points experience close encounters with the planets on short timescales, but test particles further from the Lagrange points remain for the full integration.

In our test particle survey we confirm that test particles on initially circular orbits between Jupiter and Saturn are removed by close encounters with the planets on timescales of 10^4 to 10^5 years, even when all four giant planets are included as perturbers. We also find that most test particles in the Saturn-Uranus and Uranus-Neptune regions are removed in 10 million years, with the exception of small regions

between Uranus and Neptune in which a few test particles endure for times on the order 100 million years.

Our results provide essential, new insight concerning the hypothesized Kuiper belt of comets beyond the orbit of Neptune, an expected remnant from the formation of the solar system. These are the first direct integrations that demonstrate that small bodies in low inclination, low eccentricity orbits, even as far out as 42 AU, can develop large enough eccentricities to encounter Neptune in 10 to 100 million years. As the test particles evolve into Neptune crossing orbits they roughly maintain constant semimajor axis as the eccentricity irregularly increases. Provided small bodies were formed in the Kuiper belt with nearly circular orbits, the distribution of encounter times suggests that they may only now be developing large eccentricities and being scattered by Neptune into the inner solar system.

4.8 References

Anderson, J.D. and E.M. Standish (1986). in *The Galaxy and the Solar System* ed. R. Smoluchowski, J.N. Bahcall, and M.S. Matthews (Tucson, University of Arizona) p. 286.

Arnold, V.I. (1974). *Mathematical Methods of Classical Mechanics.* Springer-Verlag, New York.

Bowell, E. (1990). *IAUC 5047.*

Bowell, E. (1990). *IAUC 5067.*

Danby, J.M.A. (1988). *Fundamental of Celestial Mechanics.* Willmann-Bell, Richmond.

Candy and Rozmus (1990). Preprint.

Cohen, C.J., E.C. Hubbard, and C. Oesterwinter (1973). *A. Papers Amer. Ephemeris* **Vol. XXII, Part 1.**

Duncan, M., T. Quinn, S. Tremaine (1988). *Astrophys. J.* **328**, L69-L73.

Duncan, M., T. Quinn, S. Tremaine (1989). *Icarus* **82**, 402-418.

Erdi, B. (1978). *C. M.* **18**, 141-161.

Erdi, B. (1979). *C. M.* **20**, 59-67.

Forest, E. and R. D. Ruth (1990). *Physica D* **43**, 105-117.

Franklin, F., M. Lecar, and P. Soper (1989). *Icarus* **79**, 223-227.

Froeschlé, Ch. and H. Scholl (1986). *Astron. Astroph.* **166**, 326-332.

Gladman, B. and M. Duncan (1990). *Astron. J.* **100**, 1680-1693.

Hamid, S.E., B.G. Marsden, and F.L. Whipple (1968). *Astron. J.* **73** 727.

Hénon, M. (1966). *Bull. Astron.* (3) 1, 57.

Heppenheimer, T.A. (1979). *Celest. Mech.* **20**, 231-241.

Hogg, D.W., G.D. Quinlan, S. Tremaine (1991). *Astron. J.* **101** 2274.

Holt, H.E. and D. Levy (1990). *IAUC 5045.*

Innanen, K. A. and S. Mikkola (1989). *Astron. J.* **97** 900-908.

Kinoshita, H., H. Yoshida, and H. Nakai (1991). *Celest. Mech.* **50**, 59-71.

Kinoshita, H. (1990). *IAUC 5075.*

Kowal, C. T. (1989). *Icarus* **77**, 118-123.

Knezevic, Z., Milani, A., Farinella, P. Froeschl'e, Ch., and Froeschl'e, Cl. (1991). *Icarus*, **93**, 316-330.

Lecar, M. and F.A. Franklin (1973). *Icarus* **20**, 422-436.

Lecar, M., F. Franklin, and M. Murison (1992). *Astron. J.* **104**, 1230-1236.

Lecar, M., F. Franklin, and P. Soper (1992). *Icarus* **96**, 234-250.

Levison, H. and M. Duncan (1990). *Astron. J.* **100**, 1669.

Levison, H. F., E. M. Shoemaker, and R. F. Wolfe (1991). *Lunar Planet. Sci. XXII*, 803-804.

Luu, J. X., and D. Jewitt (1988). *Astron. J.* **95**, 1266.

Luu, J. X., and D. Jewitt (1992). *IAUC 5611.*

Meech, K. and M. J. S. Belton (1990). *Astron. J.* **100**, 1323.

Mikkola, S. and K. Innanen (1992). *Astron. J.* **104**, 1641.

Quinn, T., Tremaine, S., and Duncan, M. (1990). *Astrophys. J.* **355** 667-679.

Shoemaker, E. M., C. S. Shoemaker, and R. F. Wolfe (1989). In **Asteroids II** (R.P Binzel, T. Gehrels, and M. S. Matthews, Eds.), pp. 487-523. University of Arizona Press, Tucson.

Soper, P., F. Franklin, and M. Lecar (1990). *Icarus* **87**, 265-284.

Sussman, G.J. and J. Wisdom (1989). *Science* **241**, 433-437.

Torbett, M. V. (1989). *Astron. J.* **98**, 1477-1481.

Torbett, M. V., a R. Smoluchowski (1990). *Nature* **345**, 49-51.

Tombaugh, C. W. (1961). In **Planets and Satellites** (G. P. Kuiper and B. M. Middlehurst, Eds.), pp. 12-30. University of Chicago Press, Chicago.

Weibel, W.M., W.M. Kaula, and W.I. Newman (1990). *Icarus* **83**, 382-390.

Wetherill, G.W. (1968). *Science* **139**, 79-82.

Wisdom, J. (1980). *Astron. J.* **85**, 1122-1133.

Wisdom, J. (1982). *Astron. J.* **87**, 577-593.

Wisdom, J. (1983). *Icarus* **56**, 51-74.

Wisdom, J. and M. Holman (1991). *Astron. J.* **102**, 1528.

Wisdom, J. and M. Holman (1992). *Astron. J.* **104**, 2022.

Yoshida, H. (1991). *Physics Letters A* **150**, 262-268.

Zhang, S.-P. and K.A. Innanen (1988). *Astron. J.* **96**, 1983.

Zhang, S.-P. and K.A. Innanen (1988). *Astron. J.* **96**, 1989.

Zhang, S.-P. and K.A. Innanen (1988). *Astron. J.* **96**, 1995.

4.9 Tables and Figures

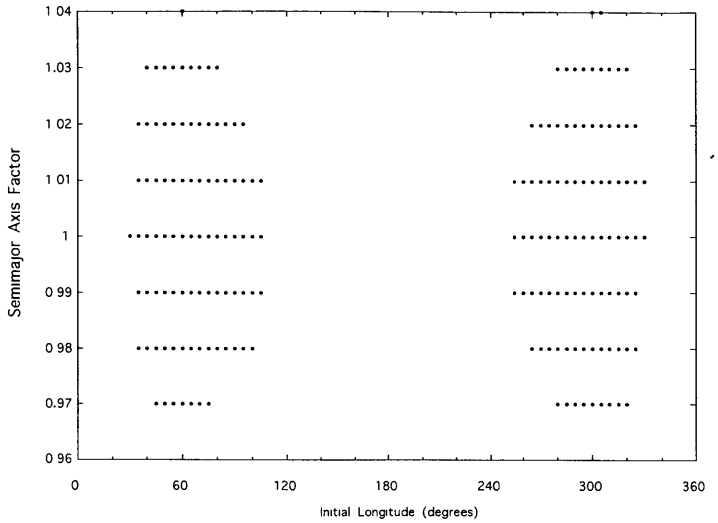
object	a (AU)	e	i (deg)	V mag	discovery	source
5335 Damocles	11.89	0.87	61.840	–	–	MPC 20799
2060 Chiron	13.74	0.38	6.93	–	Kowal	MPC 22797
5145 Pholus	20.42	0.57	24.69	–	–	MPC 19850
1993 HA2	24.80	0.523	15.63	–	Gehrels *et al.*	MPC 23350
1993 RO	32.32	–	2.53	23.2	Jewitt & Luu	IAUC 5865
1993 SB	33.15	–	2.28	23.2	Williams *et al.*	MPEC 1993-S09
1993 SC	34.45	–	5.58	22.3	Williams *et al.*	MPEC 1993-S10
1993 RP	35.37	–	2.79	25.1	Luu & Jewitt	IAUC
1992 QB1	43.83	0.088	2.21	23.6	Jewitt & Luu	IAUC 5855
1993 FW	43.93	0.041	7.75	23.4	Jewitt & Luu	IAUC 5856
1993 EV3	44.50	–	2	23.5	Jewitt & Chen	MPEC 1994-H04
1993 ES2	45.75	–	1	24.7	Luu & Jewitt	MPEC 1994-H03

Table 4.1: The preliminary designation number, semimajor axis, eccentricity, inclination, V magnitude at discovery, the discoverers, and the source for Kuiper belt candidate objects known to date.

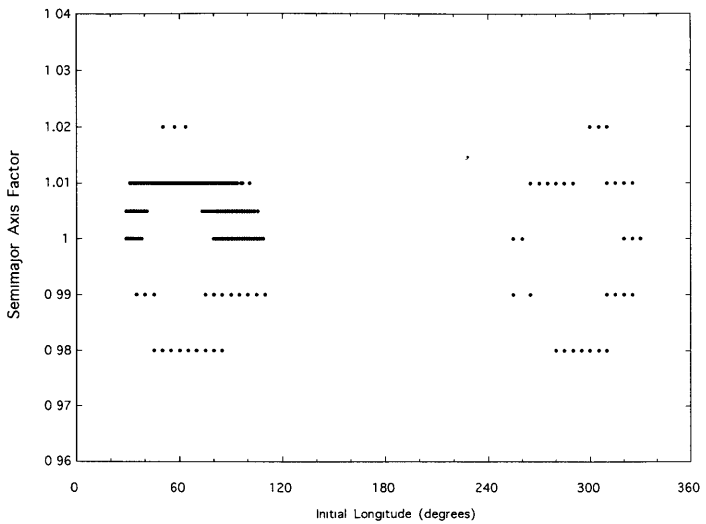
	$a_{min} - a_{max}$	Jup	Sat	Ura	Nep	$a_>$	surv	total
Jupiter	5.00-6.35	76	3	1	1	0	9	90
Jupiter-Saturn	6.35-8.24	33	92	1	0	0	0	126
Saturn	8.24-10.94	2	176	2	0	0	0	180
Saturn-Uranus	10.94-17.60	1	164	272	7	0	0	444
Uranus	17.60-20.93	0	0	215	5	0	2	222
Uranus-Neptune	20.93-27.50	0	0	195	242	0	1	438
Neptune	27.50-32.81	0	0	4	347	0	3	354
beyond Neptune	32.81-50.00	0	0	1	341	6	798	1146

Table 4.2: The number of outcomes of particles beginning in various ranges of semimajor axis. The table lists the number of encounters with each planet, the number of particles that developed semimajor axes greater than 100 AU, the number of survivors, and the total number of particles in each semimajor axis range.

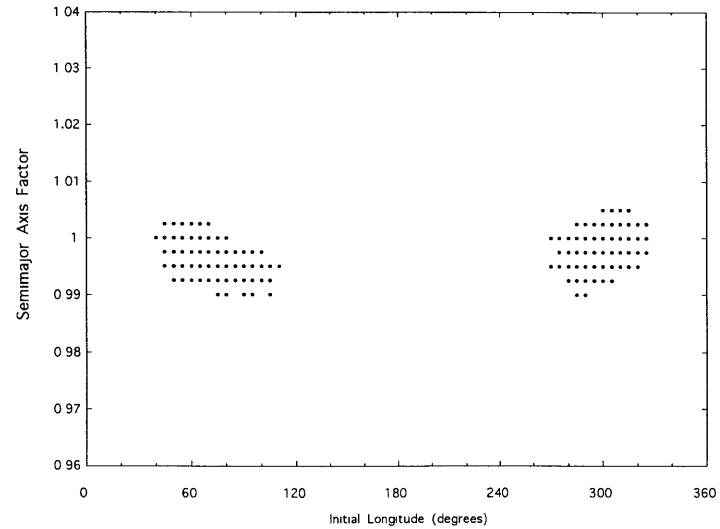
Stability Zones for Jupiter Lagrange Points,
20 million years



Stability Zones for Saturn Lagrange Points,
20 million years



Stability Zones for Uranus Lagrange Points,
20 million years



Stability Zones for Neptune Lagrange Points,
20 million years

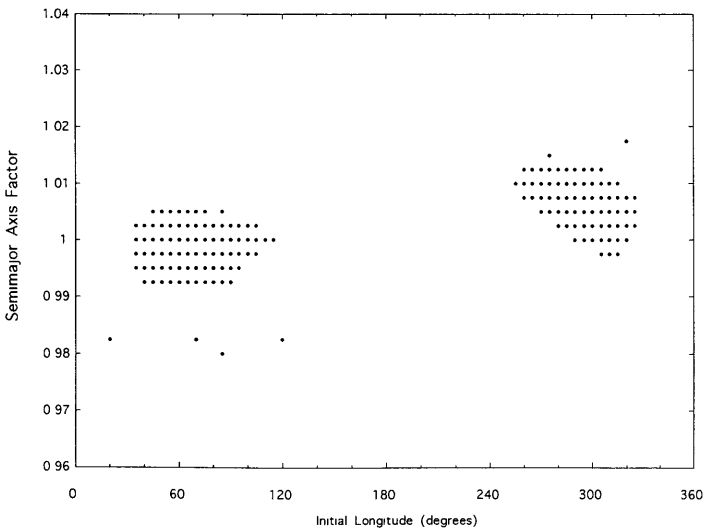


Figure 4-1: A point is plotted for each test particle that survived the full 20 million year integration. The axes show the initial displacement in longitude from the corresponding planet and factor by which that planet's semimajor axis is multiplied to initialize the semimajor axis of the test particle. A two-dimensional stable region lies near the triangular Lagrange points of each of the planets surveyed.

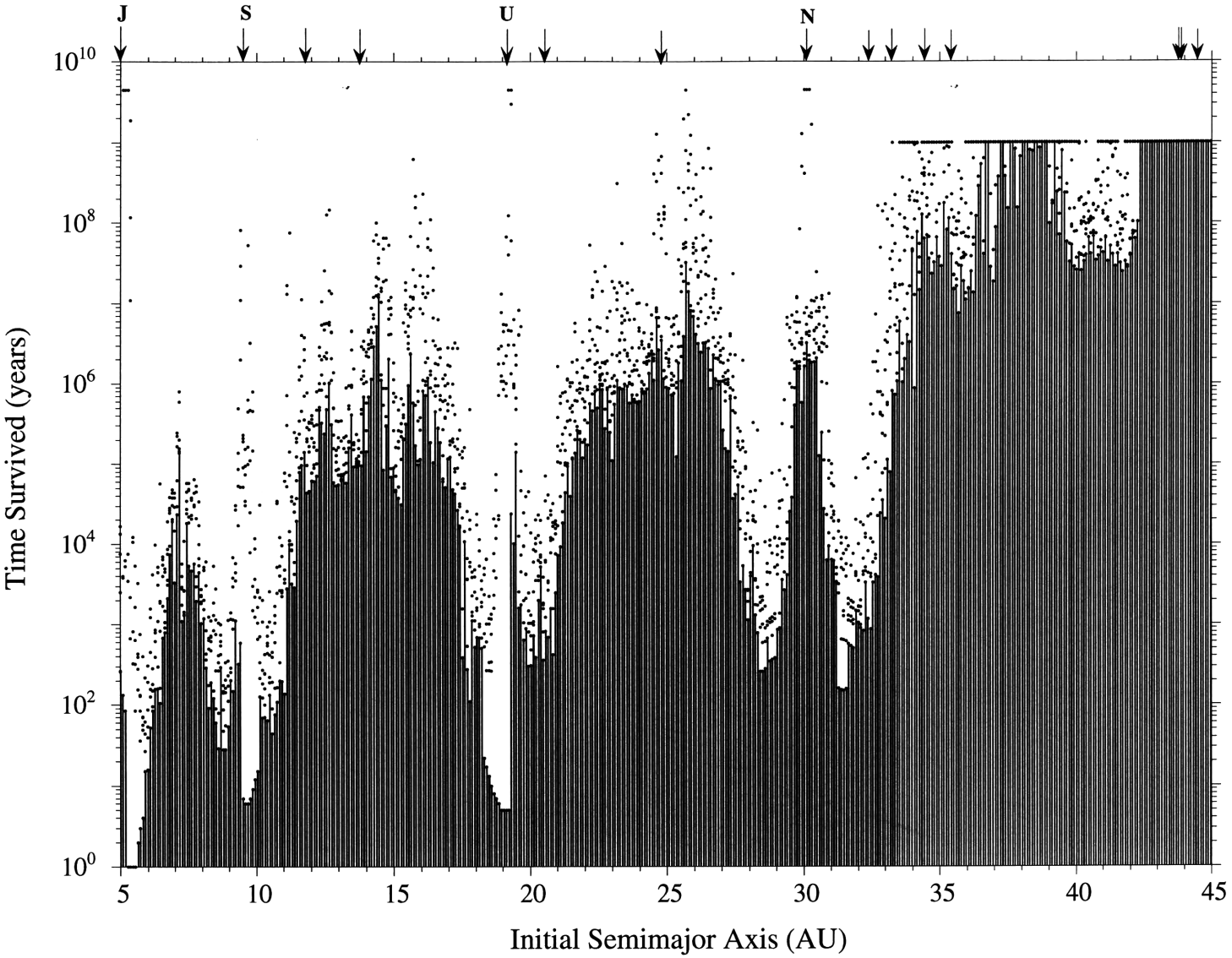

J
S
U
N
10^{10}
10^{8}
10^{6}
10^{4}
10^{2}
10^{0}
Time Survived (years)
5
10
15
20
25
30
35
40
45
Initial Semimajor Axis (AU)

Figure 4-2: The time survived by each test particle is plotted as a function of initial semimajor axis. For each semimajor axis bin, six test particles were started at different longitudes. The vertical bars mark the minimum of the six termination times. The points mark the termination time of the other five test particles. The scatter of points gives an idea of the spread of termination times for any given semimajor axis. The envelope at the top is the mark of those test particles surviving the full integration. The spikes at 5.2, 9.5, 19.2, and 30.1 AU, at the semimajor axes of the planets, correspond to test particles librating in Trojan or horseshoe-like orbits before close encounter. The semimajor axes of the planets and of other known outer solar system objects (non-comets) are marked by arrows. Interior to Neptune the integration extends to 4.5 billion years; exterior to Neptune the integration reaches 1 billion years. Beyond about 43 AU all the test particles survive the full integration.

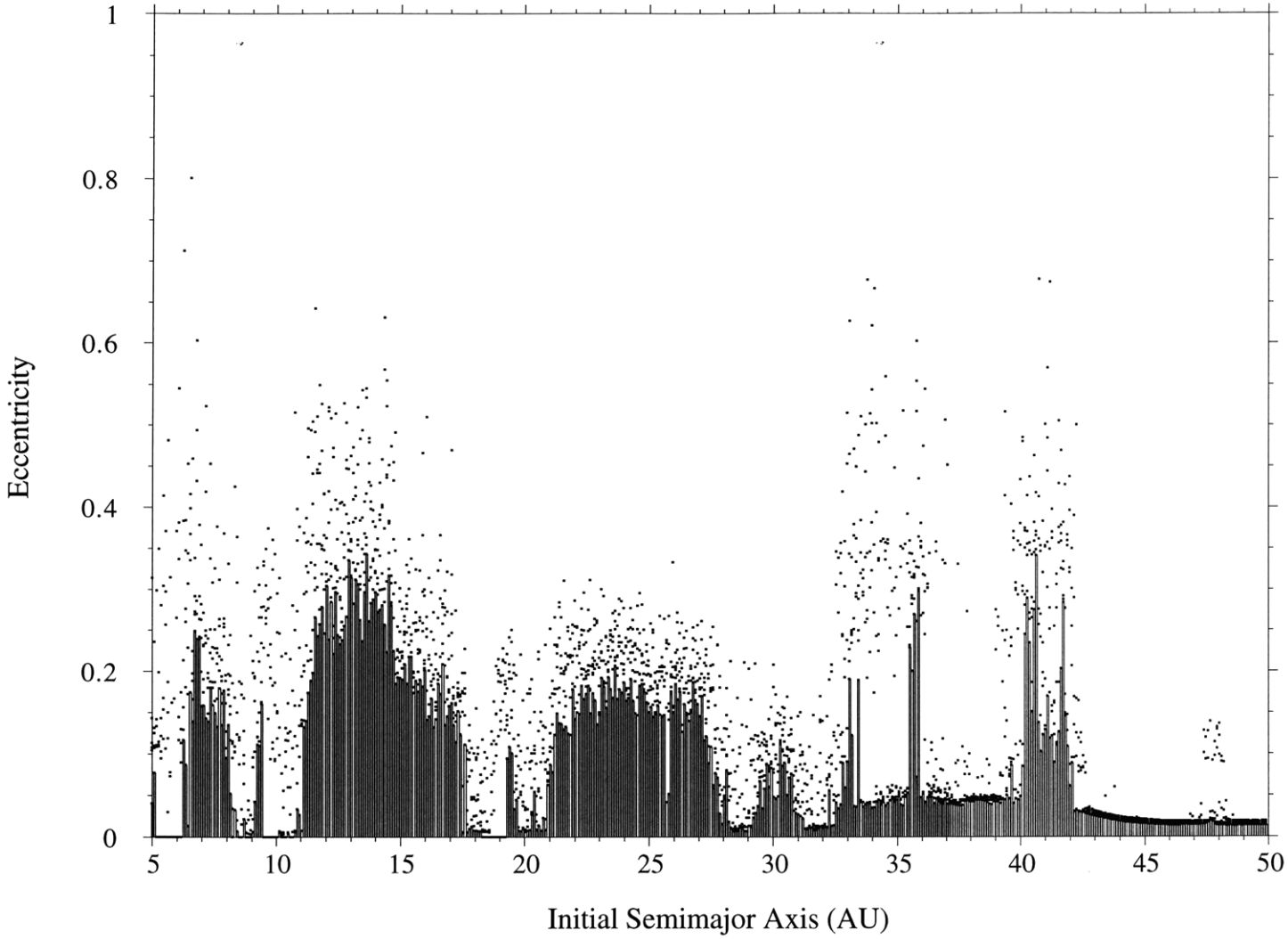

1
0.8
0.6
0.4
0.2
0
Eccentricity
5
10
15
20
25
30
35
40
45
50
Initial Semimajor Axis (AU)

Figure 4-3: The maximum eccentricity attained by each test particle during the course of the integration is plotted against initial semimajor axis. The vertical bars mark the minimum of the six values. Notice the features near 41 and 48 AU.

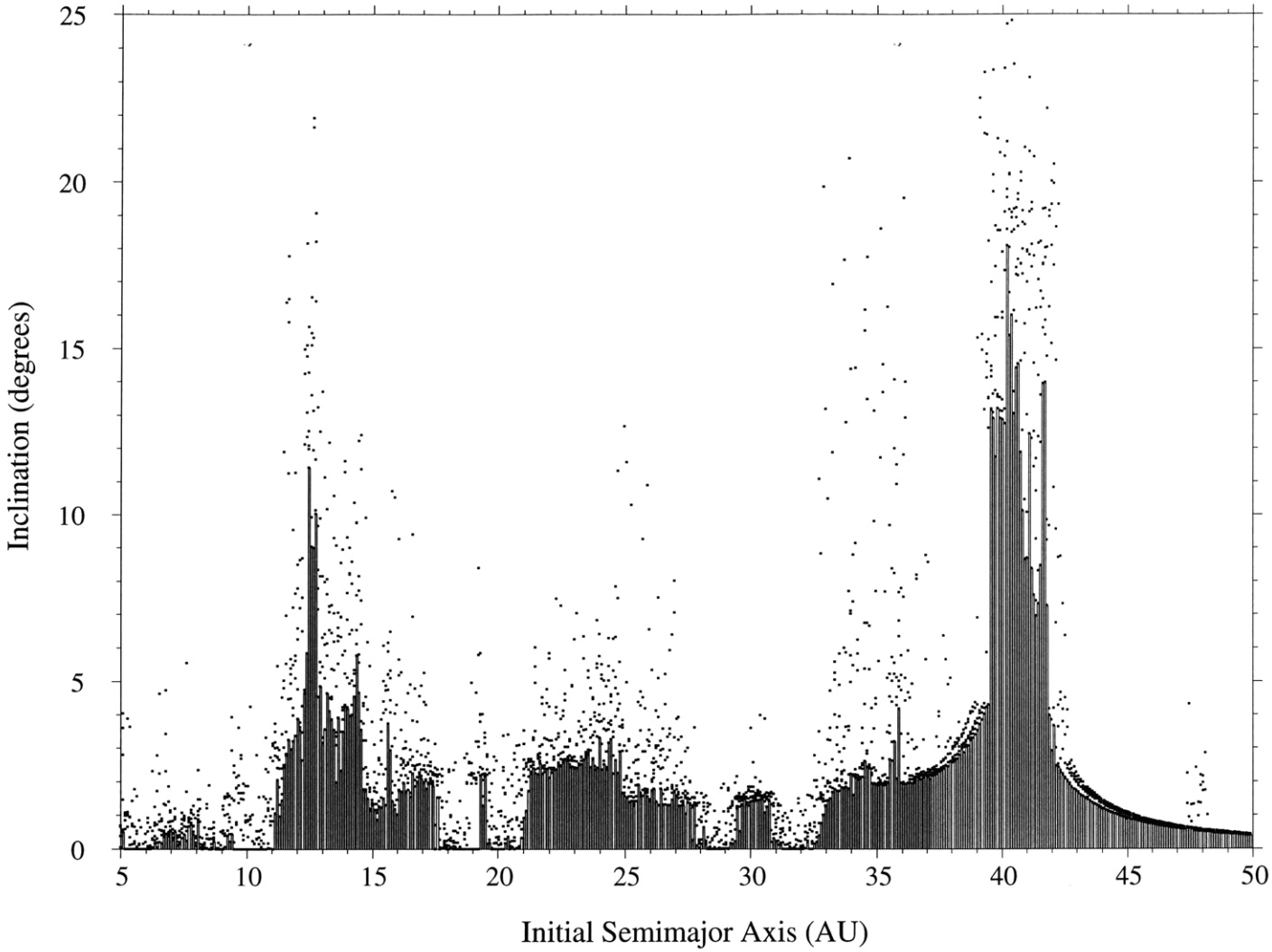

Inclination (degrees)
25
20
15
10
5
0
5
10
15
20
25
30
35
40
45
50
Initial Semimajor Axis (AU)

Figure 4-4: The maximum inclination attained by each test particle during the course of the integration is plotted against initial semimajor axis. Again, the vertical bars mark the minimum of the six values. The features near 41 and 48 AU can also be seen.

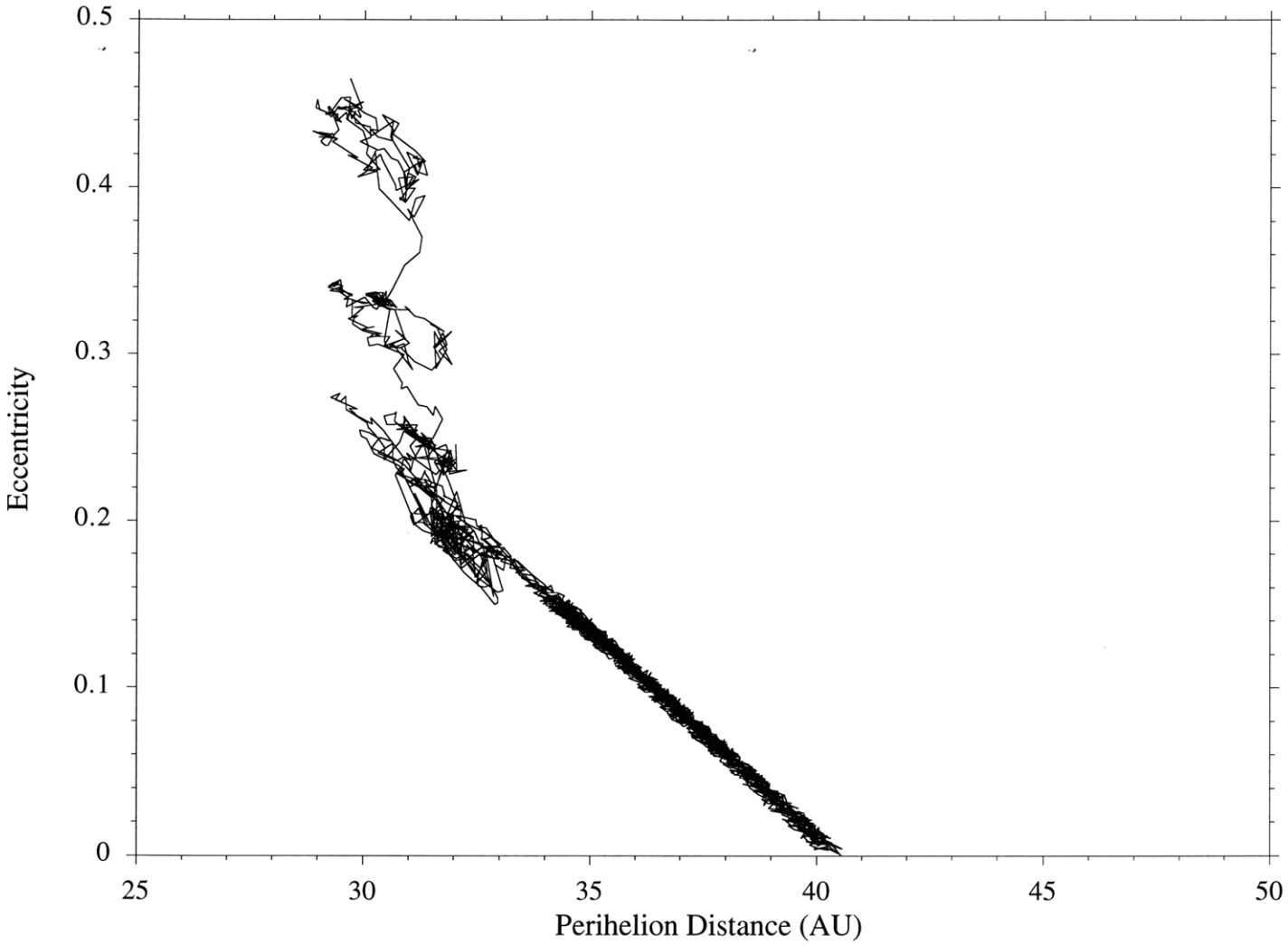

0.5
0.4
0.3
0.2
0.1
0
Eccentricity
25
30
35
40
45
50
Perihelion Distance (AU)

Figure 4-5: The eccentricity is plotted versus perihelion distance for a representative trajectory. Lines of constant semimajor axis are diagonal; lines of constant perihelion distance are vertical.

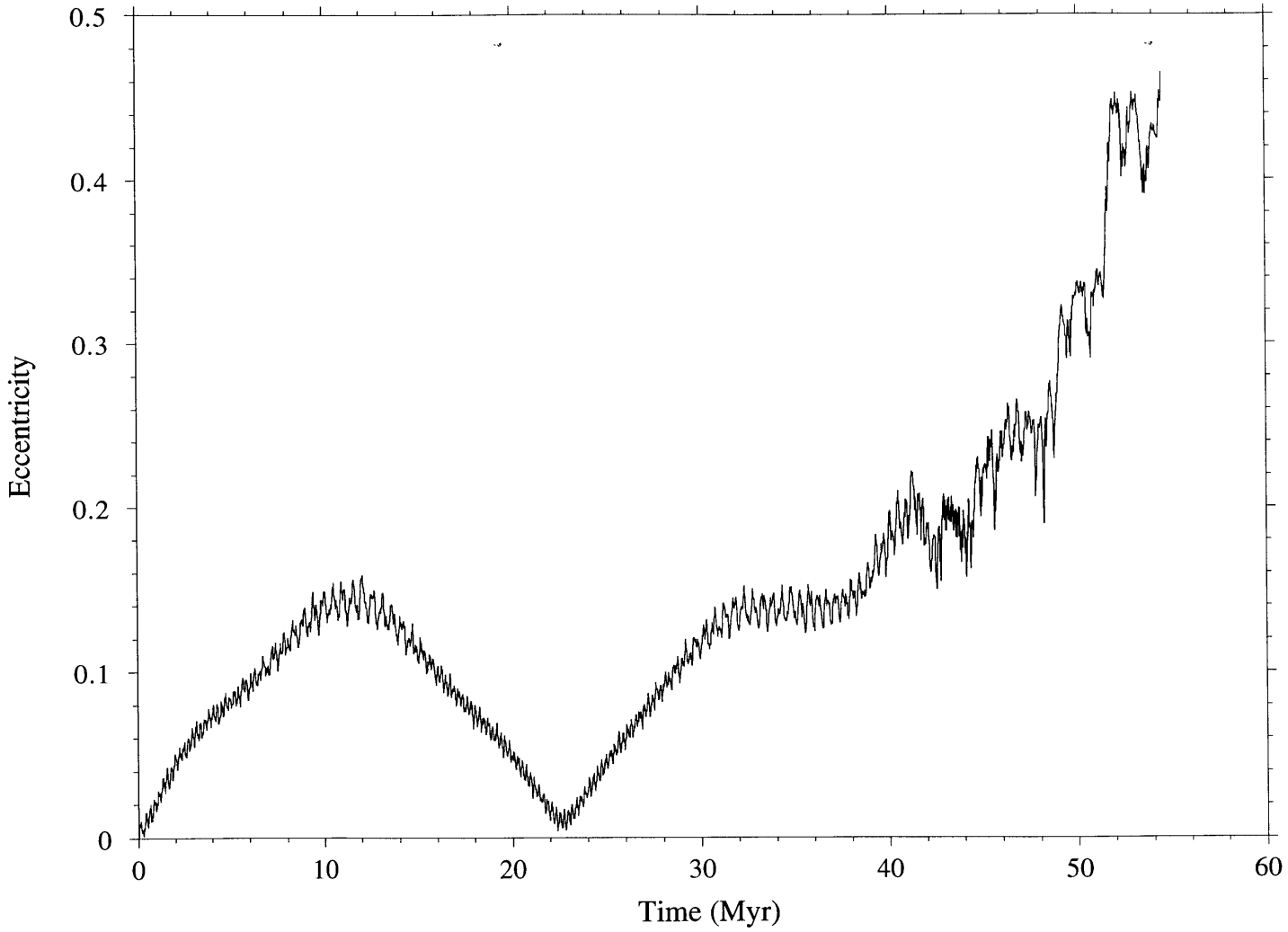

0.5
0.4
0.3
0.2
0.1
0
Eccentricity
0
10
20
30
40
50
60
Time (Myr)

Figure 4-6: Eccentricity is plotted versus time for the same trajectory as in Figure 5.

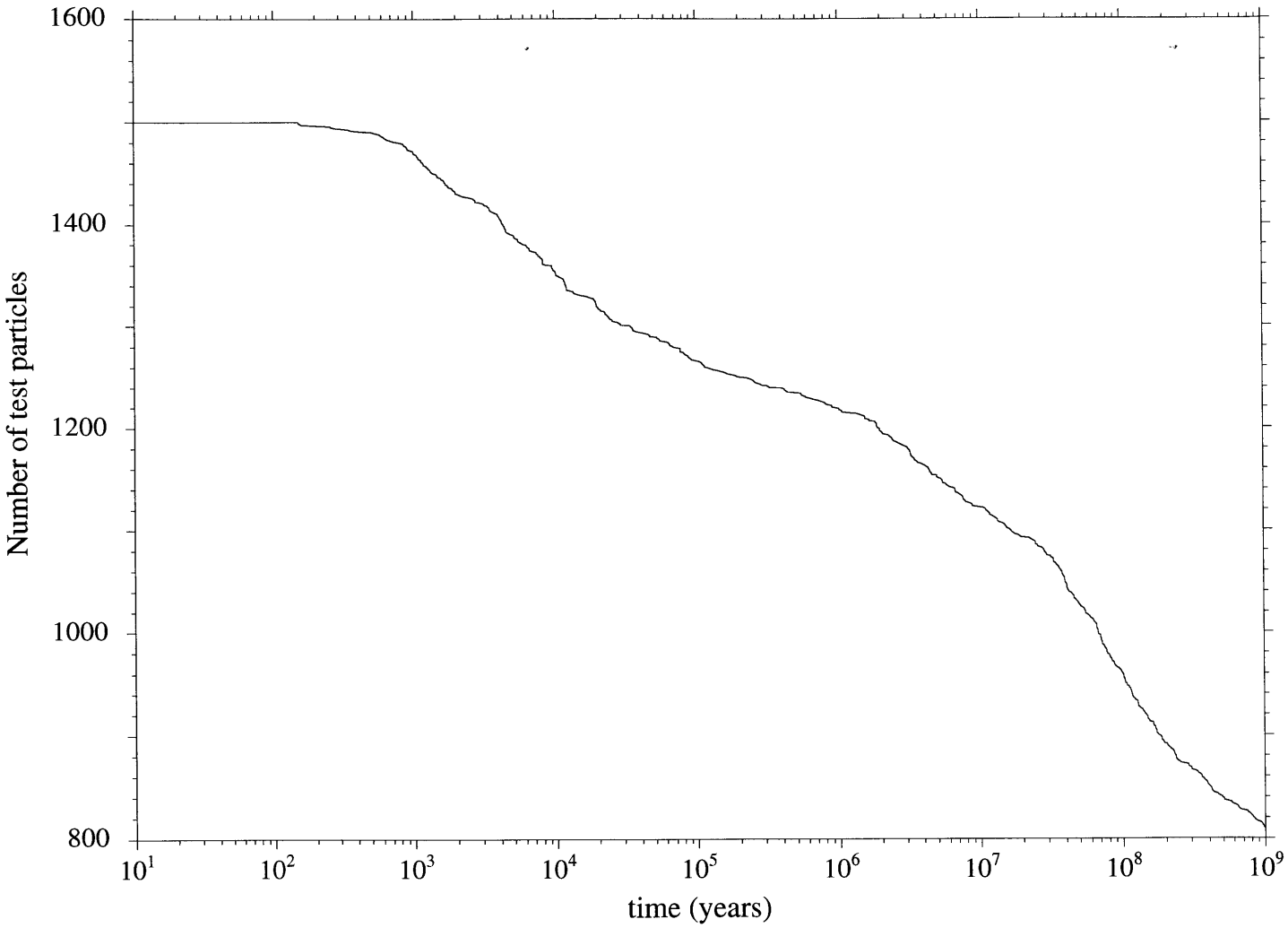

1600
1400
1200
1000
800
Number of test particles
10^1
10^2
10^3
10^4
10^5
10^6
10^7
10^8
10^9
time (years)

Figure 4-7: The number of test particles remaining beyond Neptune is plotted as a function of time. Notice the slow (logarithmic) decay.

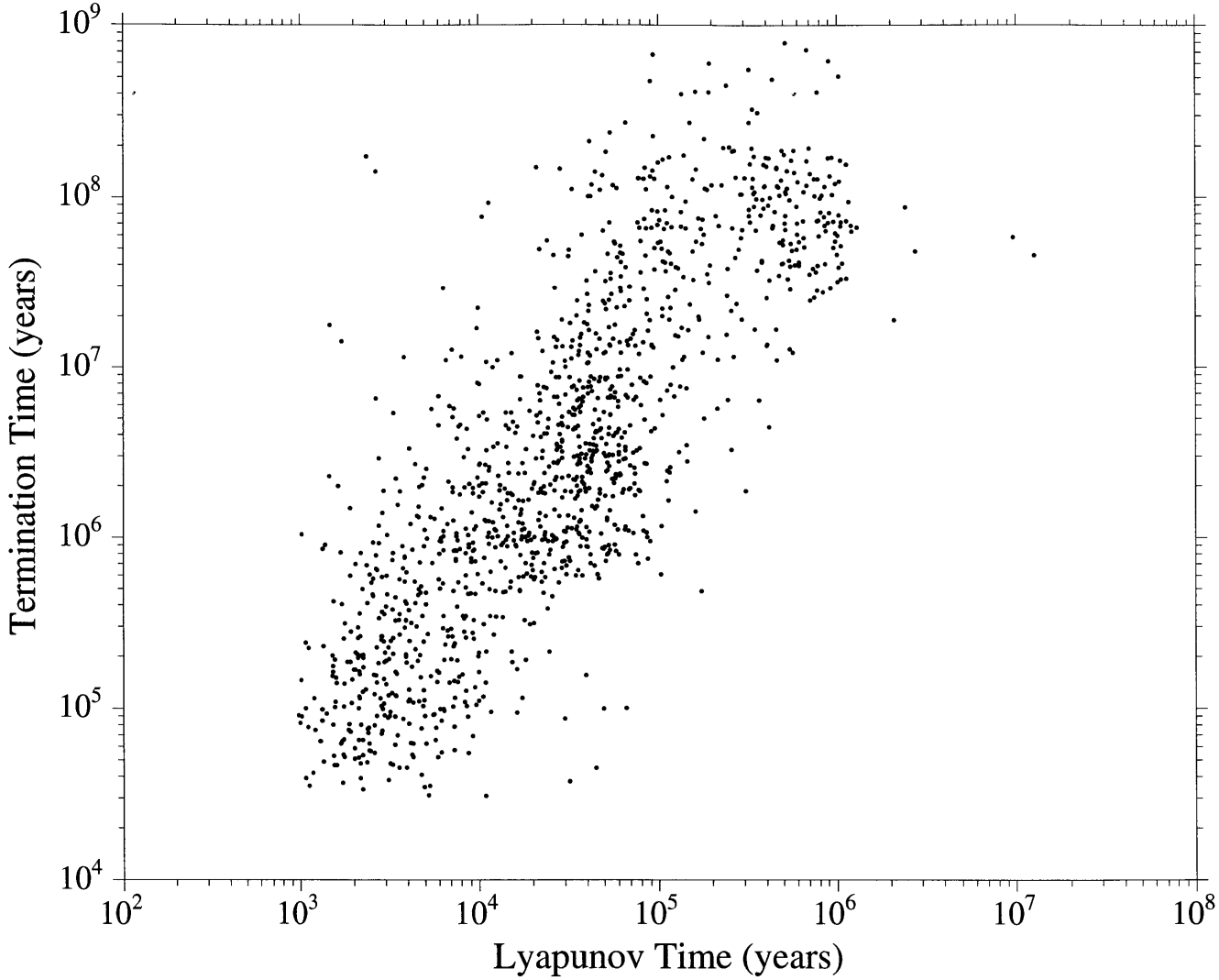

Termination Time (years)
10^9
10^8
10^7
10^6
10^5
10^4
10^2
10^3
10^4
10^5
10^6
10^7
10^8
Lyapunov Time (years)

Figure 4-8: The time of close encounter is correlated to the Lyapunov time for test particles started on circular orbits in the invariable plane.

Chapter 5

Conclusions

This thesis has presented a novel technique for integrating the solar system n-body problem. This technique is an order of magnitude faster than conventional numerical integration methods. Furthmore, the stability of the symplectic mapping method has been analyzed using the resonance overlap criterion. As an initial test, we have integrated the outer planets for 1 billion years and confirmed the result of Sussman and Wisdom (1989) that Pluto's orbit is chaotic.

Also, we have applied this technique to a survey of test particle stability in the outer solar system. On timescales of twenty million years we find no evidence that Saturn, Uranus, and Neptune cannot retain Trojan-like asteroids. Test particles in Saturn's orbit plane and placed on orbits near Saturn's L_4 and L_5 points experience close encounters with the planets on short timescales, but test particles further from the Lagrange points remain for the full integration. Also, we confirm that test particles on initially circular orbits between Jupiter and Saturn are removed by close encounters with the planets on timescales of 10^4 to 10^5 years, even when all four giant planets are included as perturbers. We also find that most test particles in the Saturn-Uranus and Uranus-Neptune regions are removed in 10 million years, with the exception of small regions between Uranus and Neptune in which a few test particles endure for times on the order 100 million years. Our results provide essential, new insight concerning the hypothesized Kuiper belt of comets beyond the orbit of Neptune, an expected remnant from the formation of the solar system. These are the

first direct integrations that demonstrate that small bodies in low inclination, low eccentricity orbits, even as far out as 42 AU, can develop large enough eccentricities to encounter Neptune in 10 to 100 million years. Provided small bodies were formed in the Kuiper belt with nearly circular orbits, the distribution of encounter times suggests that they may only now be developing large eccentricities and being scattered by Neptune into the inner solar system.

Following the research presented here, there are several directions for future work, particularly for the test particle survey of the outer solar system. First, the precise mechanisms for the delivery of short period comets from the Kuiper belt need to be examined. Although the features of a number of mean motion and secular resonances can be identified in the figures, the actual dynamics of delivery from each of these sources has not been studied in detail. Second, the integrations of test particles in the Kuiper belt region should be extended to 4.5 billion years to better determine the current flux of Kuiper belt objects encountering Neptune for the first time. Third, an effort should be made to incorporate routines for handling the passage of a test particle through a planetary close encounter into the algorithms used here. Since this sort of routine will require a variable stepsize, a transformation from mapping to real variables and its inverse must be used. This sort of transformation and its implications for the symplectic mapping method has been developed in another recent work (Wisdom, Holman, and Touma 1994). Given a routine for the passage through close encounter we can study the fate after a close encounter of the test particles in the survey of the outer solar system.