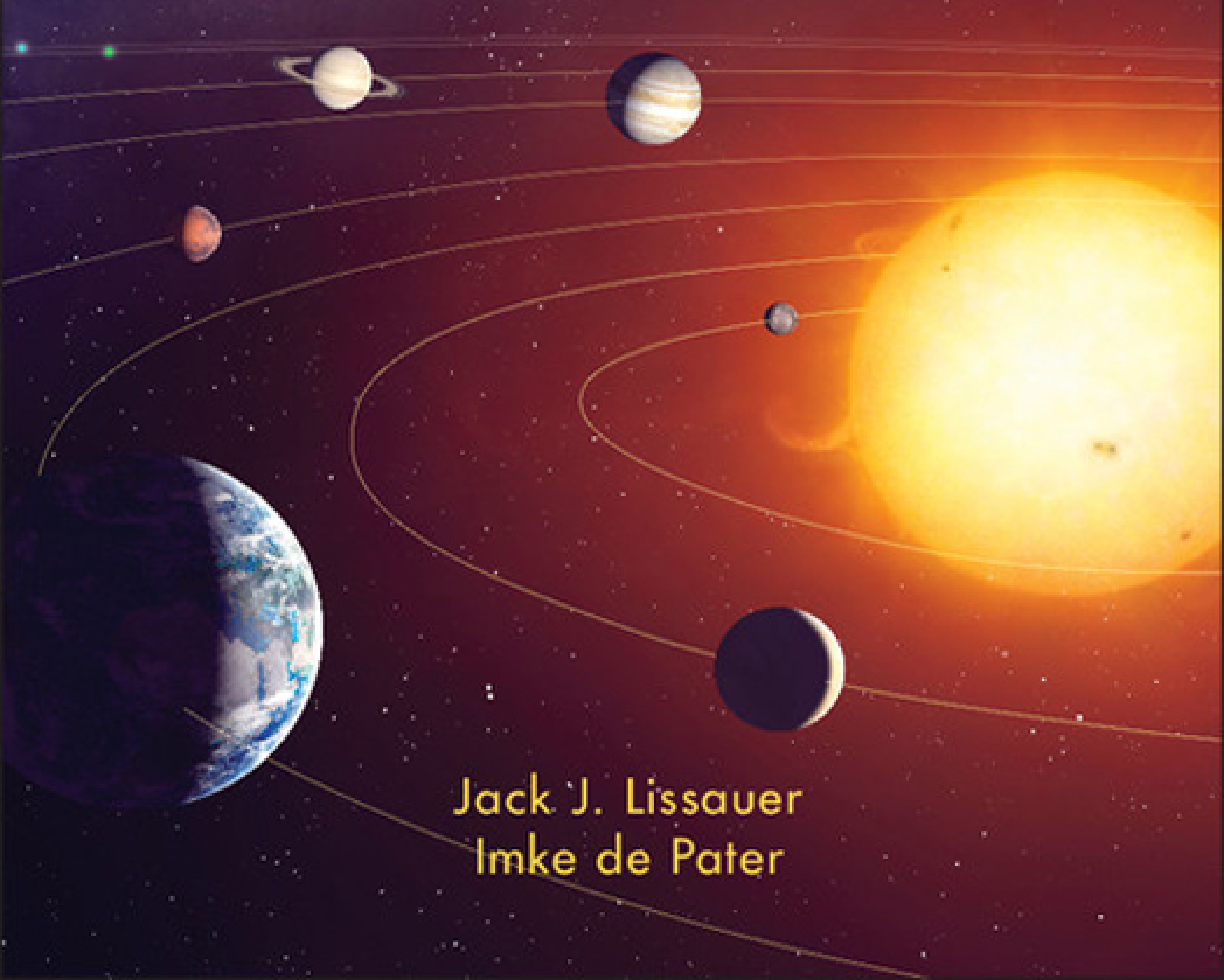
FUNDAMENTAL PLANETARY SCIENCE

Physics, Chemistry and Habitability



Jack J. Lissauer
Imke de Pater

Fundamental Planetary Science

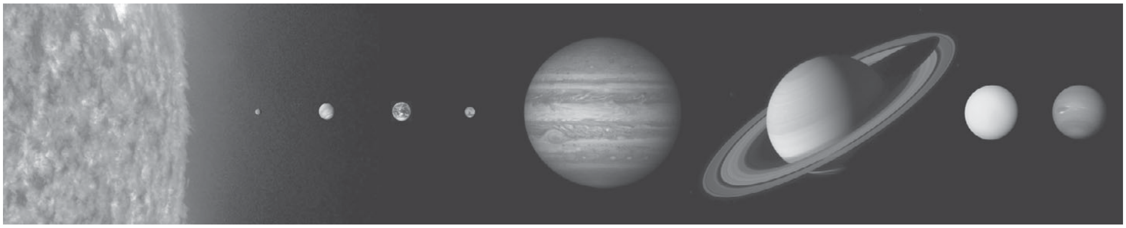
A quantitative introduction to the Solar System and planetary systems science for advanced undergraduate students, this engaging new textbook explains the wide variety of physical, chemical and geological processes that govern the motions and properties of planets, as well as how life interacts with a planet. The authors provide an overview of our current knowledge and discuss some of the unanswered questions at the forefront of research in planetary science and astrobiology today. They combine knowledge of the Solar System and the properties of extrasolar planets with astrophysical observations of ongoing star and planet formation, offering a comprehensive model for understanding the origin of planetary systems. This book concludes with an introduction to the fundamental properties of living organisms and the relationship that life has to its host planet. With more than two hundred exercises to help students learn how to apply the concepts covered, this textbook is ideal for a one-semester or two-quarter course for undergraduate students majoring in the physical or biological sciences or in engineering.

JACK J. LISSAUER is a space scientist at NASA's Ames Research Center in Moffett Field, California, and a consulting professor at Stanford University. His primary research interests are the formation of planetary systems, detection of extrasolar planets, planetary dynamics and chaos, planetary ring systems and circumstellar/protoplanetary disks. He is lead discoverer of the six-planet Kepler-11 system, and co-discoverer of the first four planets found to orbit about faint M dwarf stars; he also co-discovered two broad tenuous dust rings and two small inner moons orbiting the planet Uranus.

IMKE DE PATER is a professor in the Astronomy Department and the Department of Earth and Planetary Science at the University of California, Berkeley, and is affiliated with the Faculty of Aerospace Engineering at the Delft University of Technology, the Netherlands. She began her career observing and modeling Jupiter's synchrotron radiation followed by detailed investigations of the planet's thermal radio emission. In 1994, she led a worldwide campaign to observe the impact of comet D/Shoemaker–Levy 9 with Jupiter. Currently, she is exploiting adaptive optics techniques in the infrared range to obtain high angular resolution data of planets, their rings, and their moons in our Solar System.

Fundamental Planetary Science

PHYSICS, CHEMISTRY AND HABITABILITY



Jack J. Lissauer

NASA Ames Research Center and Stanford University

AND

Imke de Pater

University of California, Berkeley, and Delft University of Technology

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS

32 Avenue of the Americas, New York, NY 10013-2473, USA

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning and research at the highest international levels of excellence.

www.cambridge.org
Information on this title: www.cambridge.org/9780521618557

© Jack J. Lissauer and Imke de Pater 2013

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2013

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data
Lissauer, Jack Jonathan.
Fundamental planetary science : physics, chemistry and habitability / Jack Lissauer, NASA Ames Research Center and Imke de Pater, University of California, Berkeley
pages cm
Includes bibliographical references and index.
ISBN 978-0-521-85330-9 (hardback) –
ISBN 978-0-521-61855-7 (paperback)
1. Planetary theory. 2. Planets – Geology.
3. Planets – Atmospheres. 4. Planetary rings.
I. De Pater, Imke, 1952– II. Title.
QB361.L57 2013
523.9′8–dc23 2012032874

ISBN 978-0-521-85330-9 Hardback
ISBN 978-0-521-61855-7 Paperback

Additional resources and color versions of many of the illustrations are available for this publication at www.cambridge.org/lissauer.

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet websites referred to in this publication and does not guarantee that any content on such websites is, or will remain, accurate or appropriate.

CONTENTS

TABLES

PREFACE

> Astronomy compels the soul to look upwards
> and leads us from this world to another.
> Plato (427–347 BCE), *The Republic*

The wonders of the night sky, the Moon and the Sun have fascinated mankind for many millennia. We now know that objects akin to the Earth that we walk on are to be found in the heavens. What are these bodies like? What shaped them? How are they similar to our Earth, and how do they differ? And are any of them inhabited by living beings?

This text is written to provide college students majoring in the sciences with an overview of current knowledge in these areas, and the context and background to seek out and understand more detailed treatments of particular issues. We discuss what has been learned and some of the unanswered questions that remain at the forefront of planetary sciences and astrobiology research today. Topics covered include:

- the orbital, rotational and bulk properties of planets, moons and smaller bodies
- gravitational interactions, tides and resonances between bodies
- thermodynamics and other basic physics for planetary sciences
- properties of stars and formation of elements
- energy transport
- vertical structure, chemistry, dynamics and escape of planetary atmospheres
- planetary surfaces and interiors
- magnetospheres
- giant planets
- terrestrial planets
- moons
- meteorites, asteroids and comets
- planetary rings
- the new and rapidly blossoming field of extrasolar planet studies

We then combine this knowledge of current Solar System and extrasolar planet properties and processes with astrophysical data and models of ongoing star and planet formation to develop models for the origin of planetary systems. Planetary science is a key component in the new discipline of astrobiology, and a basic understanding of life is useful to planetary scientists. We therefore conclude with:

- fundamental properties of living organisms
- the relationship that life has to the planet(s) on which it forms and evolves

Parts of this book are based on the recently published second edition of our graduate textbook

Planetary Sciences. However, we have substantially modified the presentation to be more suitable for undergraduate students.

One year of calculus is required to understand all of the equations herein. Basic high school classes in physics; chemistry; and, for the final chapter of this book, biology are assumed. A college-level class designed for majors in at least one of these sciences (or in geology/geophysics or meteorology) is also expected. A small number of sections and subsections require additional background or are especially difficult; these sections are denoted with an asterisk following the section number.

The learning of concepts in the physical sciences is greatly enhanced when students get their 'hands dirty' by solving problems. Working through such exercises enables students to obtain a deeper understanding of Solar System properties. Thus, we have included an extensive collection of exercises at the end of each chapter in this text. We denote problems with a higher degree of conceptual difficulty with an asterisk.

We have used black-and-white illlustrations throughout the book, augmented with a section of color plates that repeats figures for which color is most essential to show the appearance of an object or to convey other important information. Color versions of many of the illustrations within the book are also available on the book's webpage at www.cambridge.org/lissauer. This website also includes updates, answers to selected problems (for instructors only) and links to various Solar System information sites.

Various symbols are commonly used to represent variables and constants in both equations and the text. Some variables are represented by a single standard symbol throughout the literature, and other variables are represented by differing symbols by different authors; many symbols have multiple uses. The interdisciplinary nature of the planetary sciences and astrobiology exacerbates the problem because standard notation differs between fields. We have endeavored to minimize confusion within the text and to provide the student with the greatest access to the literature by using standard symbols, sometimes augmented by nonstandard subscripts or printed using calligraphic fonts in order to avoid duplication of meanings when practical.

A list of the symbols used in this book is presented as Appendix A. Acronyms are common in our field, so we list the ones used in this book in Appendix B. Tables of physical and astronomical constants are provided in Appendix C. Appendix D is the Periodic Table of Elements. Tabulations of various properties of Solar System objects are presented in Appendix E. Because the resurgence in planetary studies during the past half century is due primarily to spacecraft sent to make close-up observations of distant bodies, we present an introduction to rocketry and list the most significant lunar and planetary missions in Appendix F. Planetary science is a rapidly advancing field, so Appendix G shows a selection of Solar System images released in 2012; we plan to update this appendix with a summary of recent developments in future printings.

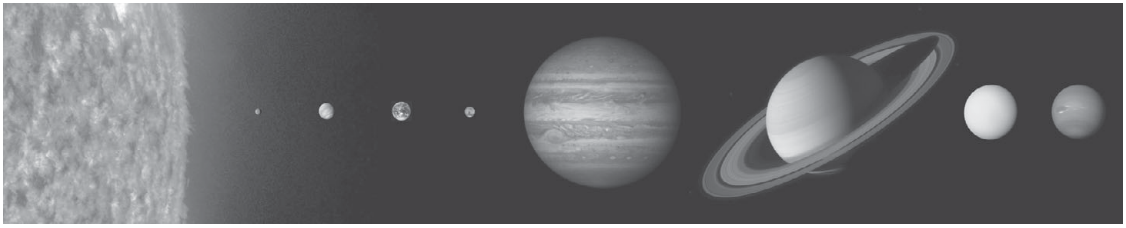
The breadth of the material covered in the text extends well beyond the areas of expertise of the authors. As such, we benefitted greatly from comments by many of our colleagues. Those who provided input for the first fifteen chapters are acknowledged in our graduate text *Planetary Sciences*, but the following group either provided significant new comments for this book or were so helpful on that text that they merit recognition here as well: Larry Esposito, Ron Greeley, Andy Ingersol, Mark Marley and Bert Vermeersen. Especially helpful suggestions for Chapter 16 were provided by Roger Linfield, Rocco Mancinelli, Frances Westall and, last but not least, Kevin Zahnle.

Jack J. Lissauer and Imke de Pater
Menlo Park, California

Fundamental Planetary Science

CHAPTER 1

Introduction



There are in fact two things, science and opinion; the former begets knowledge, the latter ignorance.

Hippocrates, *Law*, 460–377 BCE

Why are we so fascinated by planets? After all, planets make up a tiny fraction (probably substantially less than 1%) of ordinary matter in the Universe[1]. And why do terrestrial planets, which contain less than 1% of the planetary mass within our Solar System, hold a particular place in our hearts? The simple answer is that we live on a terrestrial planet. But there is a broader, more inclusive, version of that answer: To the best of our knowledge, planets or moons with solid surfaces are the only places where life can begin and evolve into advanced forms.

In this chapter, we introduce the subject of planetary sciences and provide some background needed for the remainder of the book. The history of planetary observations dates back thousands of years, and the prehistory likely extends much, much further back; we present a brief overview in the next section. We then give an inventory of objects in our Solar System in §1.2. This is followed in §1.3 by a discussion of definitions of the word 'planet' and of words describing various smaller and larger objects.

Despite the far larger number of planets known around other stars, most of our knowledge of planetary sciences was developed from observations of bodies within our own planetary system. This information is far from complete, and understanding observables is key to assessing the reliability of data; §1.4 discusses what aspects of planetary bodies we can observe.

Many lower-level planetary textbooks begin by covering the formation of our Solar System because that makes the most sense from a chronological perspective. However, although we can observe distant circumstellar disks that appear to be planetary nurseries, our observations of these disks are far less precise than those of objects orbiting the Sun. Furthermore, the accretion of planets takes a long time compared with the few decades since such observations began. Therefore, most of our understanding of planetary formation comes from a synthesis of theoretical modeling with data from our own Solar System and extrasolar planets. We thus defer our main discussion of this subject, which is among the most intellectually challenging in planetary science, until near the end of this book. Nonetheless, scientists have modeled the origin of planets for hundreds of years, and our understandings of this process have provided the best estimates of certain planetary properties that are not directly observable, such as interior composition. Because interpretation of data and planetary formation models often go hand in hand, we present a brief summary of current models of planetary formation in the final section of this chapter.

[1] **Dark matter**, most of which is **nonbaryonic** (i.e., not composed of protons or neutrons), is an order of magnitude more abundant than ordinary matter, which is also referred to as **luminous matter**. **Dark energy** has more than twice the mass-energy density of all types of matter in the Universe combined.

1.1 A Brief History of the Planetary Sciences

The sky appears quite spectacular on a clear night away from the light of modern cities. Ancient civilizations were particularly intrigued by several brilliant 'stars' that move among the far more numerous 'fixed' (stationary) stars. The Greeks called these objects **planets**, or wandering stars. Old drawings and manuscripts by people from all over the world, including the Chinese, Greeks and Anasazi, attest to their interest in comets, solar eclipses and other celestial phenomena. And observations of planets surely date to well before the dawn of writing and historical records, perhaps predating humanity itself. Some migratory birds use the patterns of stars in the night sky to guide their journeys and might be aware that a few of these objects move relative to the others. Indeed, some sharp-eyed and keen-witted dinosaurs may have realized that a few points of light in the night sky moved relative to the fixed pattern produced by most ‘stars’ more than 100 million years ago, but as dinosaurs never (to our knowledge) developed a written language, it is unlikely that such speculation will ever be confirmed.

The Copernican–Keplerian–Galilean–Newtonian revolution in the sixteenth and seventeenth centuries completely changed humanity’s view of the dimensions and dynamics of the Solar System, including the relative sizes and masses of the bodies and the forces that make them orbit about one another. Gradual progress was made over the next few centuries, but the next revolution had to await the space age.

The age of planetary exploration began in October of 1959, with the Soviet Union’s spacecraft *Luna 3* returning the first pictures of the farside of Earth’s Moon (Fig. F.1). Over the next three decades, spacecraft visited all eight known terrestrial and giant planets in the Solar System, including our own. These spacecraft have returned data concerning the planets, their rings and moons. Spacecraft images of many objects showed details never suspected from earlier Earth-based pictures. Spectra from γ-rays to radio wavelengths revealed previously undetected gases and geological features on planets and moons, and radio detectors and magnetometers transected the giant magnetic fields surrounding many of the planets. The planets and their satellites have become familiar to us as individual bodies. The immense diversity of planetary and satellite surfaces, atmospheres and magnetic fields has surprised even the most imaginative researchers. Unexpected types of structure were observed in Saturn’s rings, and whole new classes of rings and ring systems were seen around all four giant planets. Some of the new discoveries have been explained, but others remain mysterious.

Five comets and ten asteroids have thus far been explored close up by spacecraft (Table F.2), and there have been several missions to study the Sun and the solar wind. The Sun’s gravitational domain extends thousands of times the distance to the farthest known planet, Neptune. Yet the vast outer regions of the Solar System are so poorly explored that many bodies remain to be detected, possibly including some of planetary size.

Hundreds of planets are now known to orbit stars other than the Sun. Although we know far less about any of these extrasolar planets than we do about the planets in our Solar System, it is clear that many of them have gross properties (orbits, masses, radii) quite different from any object orbiting our Sun, and they are thus causing us to revise some of our models of how planets form and evolve.

Biologists have redrawn the tree of life over the past few decades. We have learned of the interrelationships between all forms of life on Earth and of life’s great diversity. This diversity enables some species to live in environments that would be considered quite extreme to humans and suggests that conditions capable of sustaining life exist on other planets and moons in our Solar System and beyond.

The renewed importance of the planetary sciences as a subfield of astronomy implies that some exposure to Solar System studies is an important component to the education of astronomers. Planetary sciences’ close relationship to geophysics, atmospheric and space sciences means that the study of the planets offers the unique opportunity for comparison available to Earth scientists. The properties of planets are key to astrobiology, and understanding the basics of life is useful to planetary scientists.

1.2 Inventory of the Solar System

What is the **Solar System**? Our naturally geocentric view gives a highly distorted picture; thus, it is better to phrase the question as: What is seen by an objective observer from afar? The **Sun**, of course; the Sun has a luminosity 4×10^8 times as large as the total luminosity (reflected plus emitted) of Jupiter, the second brightest object in the Solar

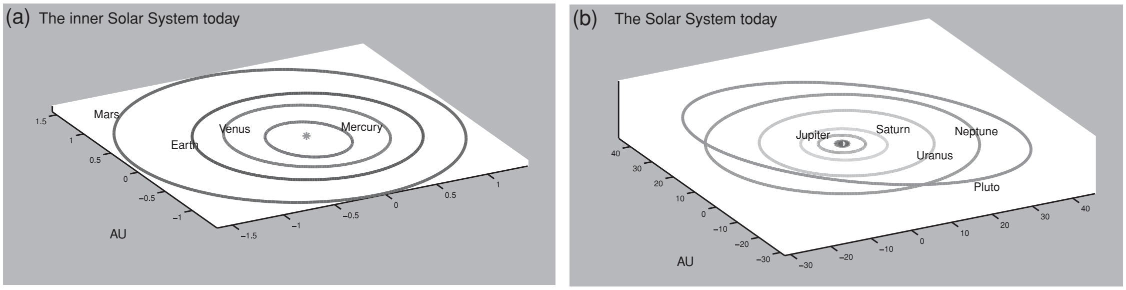


Figure 1.1 The orbits of (a) the four terrestrial planets and (b) all eight major planets in the Solar System and Pluto, are shown to scale. The axes are in AU. The movies show variations in the orbits over the past 3 million years; these changes are caused by mutual perturbations among the planets (see Chapter 2). Figure 2.12 presents plots of the variations in planetary eccentricities from the same integrations. (Illustrations courtesy Jonathan Levine)

System. The Sun also contains >99.8% of the mass of the known Solar System. By these measures, the Solar System can be thought of as the Sun plus some debris. However, by other measures, the planets are not insignificant. More than 98% of the angular momentum in the Solar System lies in orbital motions of the planets. Moreover, the Sun is a fundamentally different type of body from the planets – a ball of plasma powered by nuclear fusion in its core – but the smaller bodies in the Solar System are composed of molecular matter, some of which is in the solid state. This book focusses on the debris in orbit about the Sun, although we do include a summary of the properties of stars, including our Sun, in §3.3, and an overview of the outer layers of the Sun and its effect on the interplanetary medium in §§7.1 and 7.2. The debris encircling the Sun is composed of the giant planets, the terrestrial planets and numerous and varied smaller objects.

Figures 1.1 to 1.3 present three differing views of the Solar System. The orbits of the major planets and Pluto are diagrammed in Figure 1.1. Two different levels of reduction are displayed because of the relative closeness of the four terrestrial planets and the much larger spacings in the outer Solar System. Note the high inclination of Pluto's orbit relative to the orbits of the major planets. Figure 1.2 plots the sizes of various classes of Solar System objects as a function of location. The jovian (giant) planets dominate the outer Solar System, and the terrestrial planets dominate the inner Solar System. Small objects tend to be concentrated in regions where orbits are stable or at least long lived. Images of the planets and the largest planetary satellites are presented to scale in Figure 1.3. Figure 1.4 shows close-up views of those comets and asteroids that had been imaged by interplanetary spacecraft as of 2010.

1.2.1 Giant Planets

Jupiter dominates our planetary system. Its mass, 318 Earth masses ($M_\oplus$), exceeds twice that of all other known Solar System planets combined. Thus, as a second approximation, the Solar System can be viewed as the Sun, Jupiter and some debris. The largest of this debris is **Saturn**, with a mass of nearly 100 $M_\oplus$. Saturn, similar to Jupiter, is made mostly of hydrogen (H) and helium (He). Each of these planets probably possesses a heavy element 'core' of mass ~10 $M_\oplus$. The third and fourth largest planets are **Neptune** and **Uranus**, each having a mass roughly one-sixth that of Saturn. These planets belong to a different class, with most of their masses provided by a combination of three

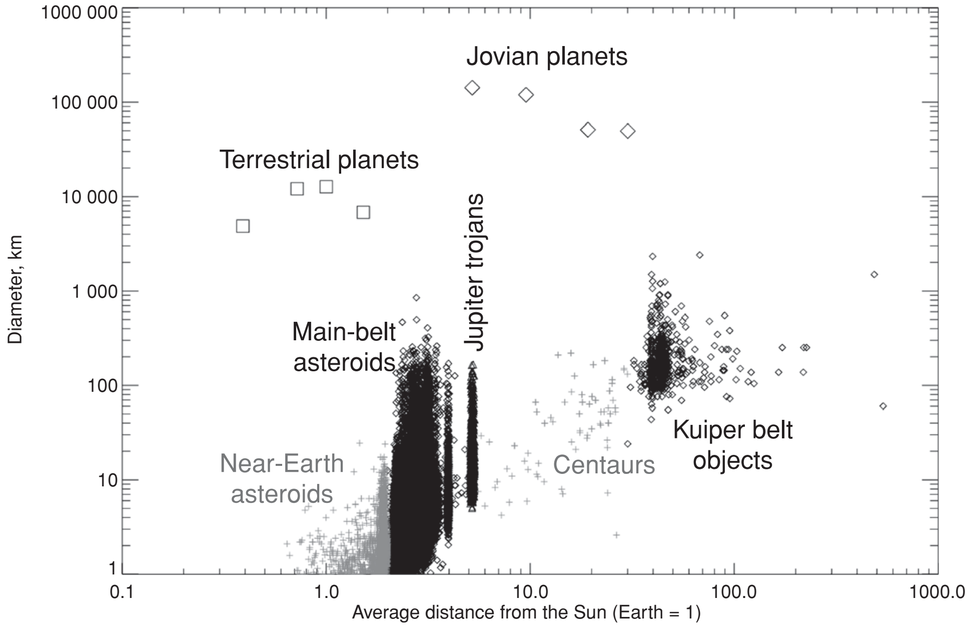


Figure 1.2 Inventory of objects orbiting the Sun. Small bodies are discussed in Chapter 12. The orbits of Jupiter Trojans are described in §2.2.1 and those of Centaurs are discussed in §12.2.2. (Courtesy John Spencer)

common astrophysical 'ices', water (H_2O), ammonia (NH_3), methane (CH_4), together with 'rock', high temperature condensates consisting primarily of silicates and metals, yet most of their volumes are occupied by relatively low mass ($1–4\ M_\oplus$) H–He dominated atmospheres. The four largest planets are known collectively as the **giant planets**; Jupiter and Saturn are called **gas giants**, with radii of $\sim$70 000 km and 60 000 km, respectively, and Uranus and Neptune are referred to as **ice giants** (although the 'ices' are present in fluid rather than solid form), with radii of $\sim$25 000 km. All four giant planets possess strong magnetic fields. These planets orbit the Sun at distances of approximately 5, 10, 20 and 30 AU, respectively. (One **astronomical unit**, 1 AU, is defined to be the semimajor axis of a massless [test] particle whose orbital period about the Sun is one year. As our planet has a finite mass, the semimajor axis of Earth's orbit is slightly larger than 1 AU.)

1.2.2 Terrestrial Planets

The mass of the remaining known 'debris' totals less than one-fifth that of the smallest giant planet, and their orbital angular momenta are also much smaller. This debris consists of all of the solid bodies in the Solar System, and despite its small mass, it contains a wide variety of objects that are interesting chemically, geologically, dynamically, and, in at least one case, biologically. The hierarchy continues within this group, with two large **terrestrial**[2] planets, **Earth** and **Venus**, each with a radius of about 6000 km, at approximately 1 and 0.7 AU from the Sun, respectively. Our Solar System also contains two small terrestrial planets,

[2] In this text, the word 'terrestrial' is used to mean Earth-like or related to the planet Earth, as is the convention in planetary sciences and astronomy. Geoscientists and biologists generally use the same word to signify a relationship with land masses.

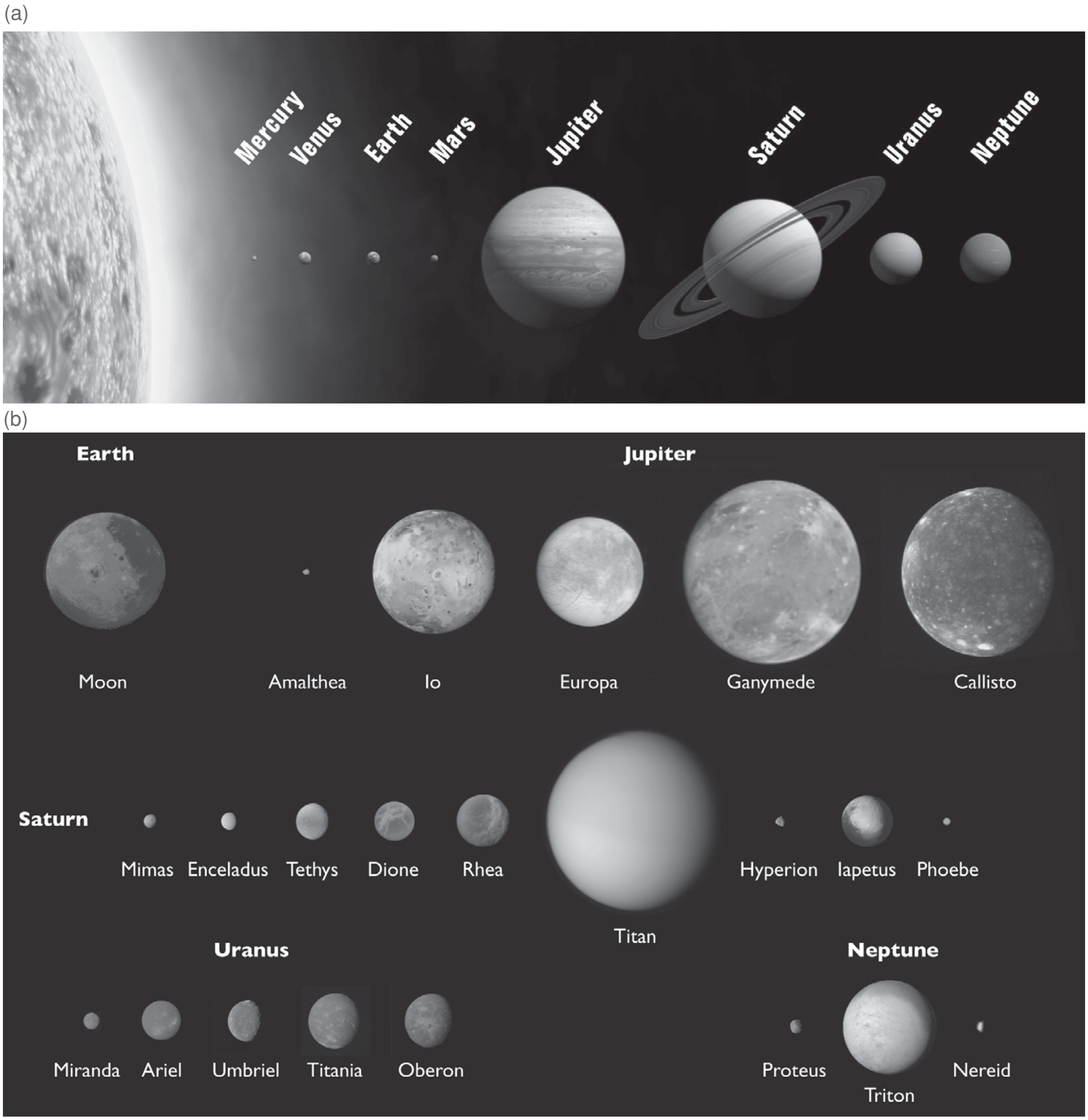


Figure 1.3 COLOR PLATE (a) Images of the planets with radii depicted to scale, ordered by distance from the Sun. (Courtesy International Astronomical Union/Martin Kornmesser) (b) Images of the largest satellites of the four giant planets and Earth's Moon, which are depicted in order of distance from their planet. Note that these moons span a wide range of size, albedo (reflectivity) and surface characteristics; most are spherical, but some of the smallest objects pictured are quite irregular in shape. (Courtesy Paul Schenk)

Mars with a radius of ~3500 km and orbiting at ~1.5 AU and **Mercury** with a radius of ~2500 km orbiting at ~0.4 AU.

All four terrestrial planets have atmospheres. Atmospheric composition and density vary widely among the terrestrial planets, with Mercury's atmosphere being exceedingly thin. However, even the most massive terrestrial planet atmosphere, that of Venus, is minuscule by giant planet standards. Earth and Mercury each have an internally

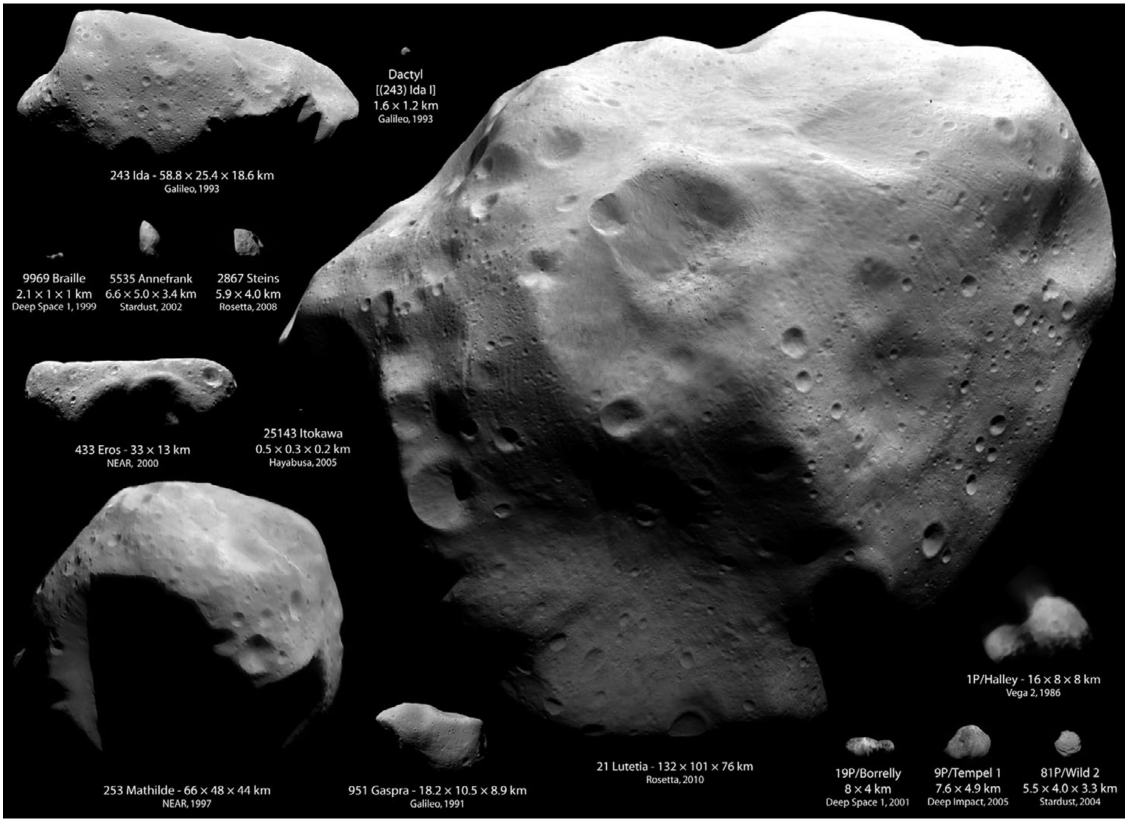


Figure 1.4 Views of the first four comets (lower right) and nine asteroid systems that were imaged close-up by interplanetary spacecraft, shown at the same scale. The object name and dimensions, as well as the name of the imaging spacecraft and the year of the encounter, are listed below each image. Note the wide range of sizes. Dactyl is a moon of Ida.

generated magnetic field, and evidence suggests that Mars possessed one in the distant past.

1.2.3 Minor Planets and Comets

The **Kuiper belt** is a thick disk of ice/rock bodies beyond the orbit of Neptune. The two largest members of the Kuiper belt to have been sighted are **Eris**, whose **heliocentric distance**, the distance from the Sun, oscillates between 38 and 97 AU, and **Pluto**, whose heliocentric distance varies from 29 to 50 AU. The radii of Eris and Pluto exceed 1000 km. Pluto is known to possess an atmosphere. Numerous smaller members of the Kuiper belt have been cataloged, but the census of these distant objects is incomplete even at large sizes. **Asteroids**, which are minor planets that all have radii <500 km, are found primarily between the orbits of Mars and Jupiter.

Smaller objects are also known to exist elsewhere in the Solar System, for example as moons in orbit around planets, and as comets. Comets are ice-rich objects that shed mass when subjected to sufficient solar heating. Comets are thought to have formed in or near the giant planet region and then been 'stored' in the **Oort cloud**, a nearly spherical region at heliocentric distances of $\sim 1-5 \times 10^4$ AU, or in the Kuiper belt or the **scattered disk**. Scattered disk objects (SDOs) have moderate to high eccentricity orbits that lie in whole or in part within the Kuiper belt. Estimates of the total number of comets larger than 1 km in radius in the entire Oort cloud range from $\sim 10^{12}$ to $\sim 10^{13}$. The total number of Kuiper belt objects (KBO) larger than 1 km in radius is estimated to be $\sim 10^{8}-10^{10}$. The total mass and orbital angular momentum of bodies in the scattered disk and Oort cloud are uncertain by more than an order of magnitude. The upper end of current estimates place as much mass in distant unseen icy bodies as is observed in the entire planetary system.

The smallest bodies known to orbit the Sun, such as the dust grains that together produce the faint band in the plane of the planetary orbits known as the **zodiacal cloud**, have been observed collectively but not yet individually detected via remote sensing.

1.2.4 Satellite and Ring Systems

Some of the most interesting objects in the Solar System orbit about the planets. Following the terrestrial planets in mass are the seven major moons of the giant planets and Earth. Two planetary satellites, Jupiter's moon Ganymede and Saturn's moon Titan, are slightly larger than the planet Mercury, but because of their lower densities, they are less than half as massive. Titan's atmosphere is denser than that of Earth. Triton, by far the largest moon of Neptune, has an atmosphere that is much less dense, yet it has winds powerful enough to strongly perturb the paths of particles ejected from geysers on its surface. Very tenuous atmospheres have been detected about several other planetary satellites, including Earth's Moon, Jupiter's Io and Saturn's Enceladus.

Natural satellites have been observed in orbit about most of the planets in the Solar System, as well as many Kuiper belt objects and asteroids. The giant planets all have large satellite systems, consisting of large- and/or medium-sized satellites (Fig. 1.3b) and many smaller moons and rings. Most of the smaller moons orbiting close to their planet were discovered from spacecraft flybys. All major satellites, except Triton, orbit the respective planet in a **prograde** manner (i.e., in the direction that the planet rotates) close to the planet's equatorial plane. Small, close-in moons are also exclusively in low-inclination, low-eccentricity orbits, but small moons orbiting beyond the main satellite systems can travel around the planet in either direction, and their orbits are often highly inclined and eccentric. Earth and Pluto each have one large moon: our Moon has a little over 1% of Earth's mass, and Charon's mass is just over 10% that of Pluto. These moons probably were produced by giant impacts on the Earth and Pluto when the Solar System was a small fraction of its current age. Two tiny moons travel on low-inclination, low-eccentricity orbits about Mars.

The four giant planets all have ring systems, which are primarily located within about 2.5 planetary radii of the planet's center. However, in other respects, the characters of the four ring systems differ greatly. Saturn's rings are bright and broad, full of structure such as density waves, gaps and 'spokes'. Jupiter's ring is very tenuous and composed mostly of small particles. Uranus has nine narrow opaque rings plus broad regions of tenuous dust orbiting close to the plane defined by the planet's equator. Neptune has four rings, two narrow ones and two faint broader rings; the most remarkable part of Neptune's ring system is the ring arcs, which are bright segments within one of the narrow rings.

1.2.5 Tabulations

The orbital and bulk properties of the eight 'major' planets are listed in Tables E.1 to E.3. Symbols for each of these planets, which we often use as subscripts on masses and radii, are also given in Table E.1. Table E.4 gives orbital elements and brightnesses of all inner moons of the eight planets, as well as those outer moons whose radii are estimated to be $\gtrsim 10$ km. Many of the orbital parameters listed in the tables are defined in §2.1. Rotation rates and physical characteristics of these satellites, whenever known, are given in Table E.5. Properties of some the largest 'minor planets',

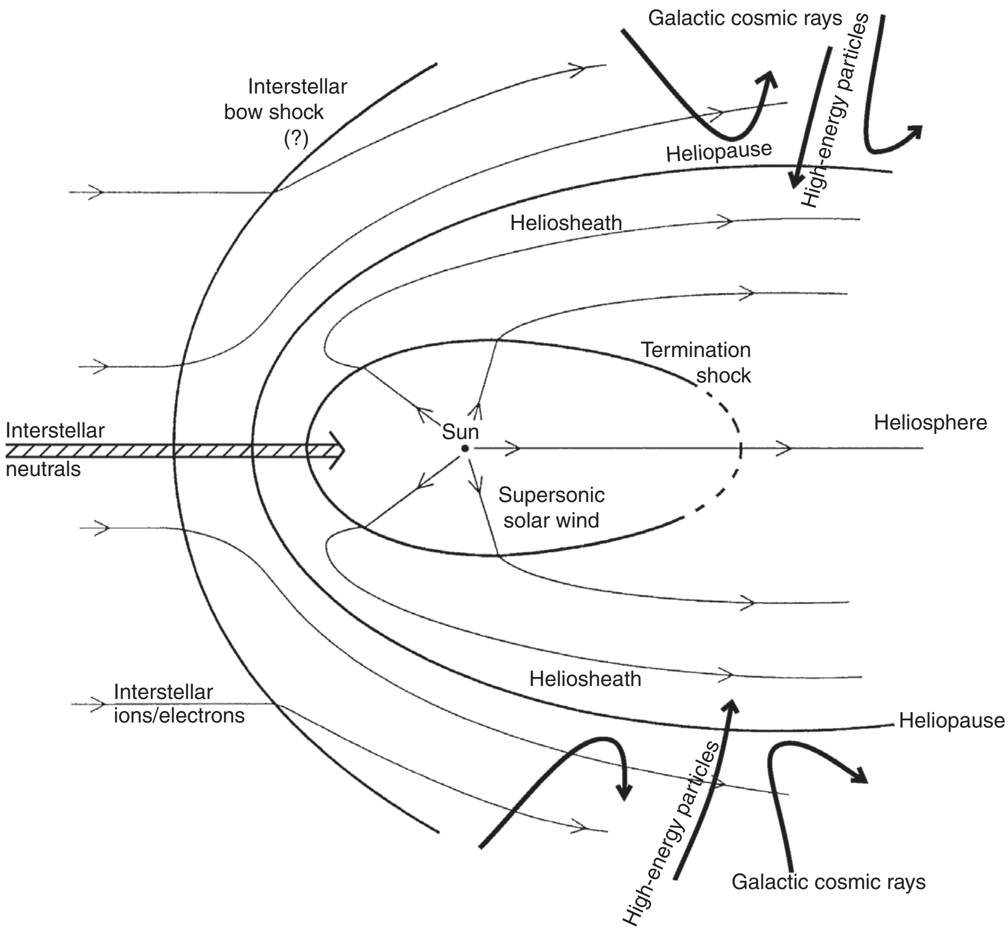


Figure 1.5 Sketch of the teardrop-shaped heliosphere. Within the heliosphere, the solar wind flows radially outwards until it encounters the heliopause, the boundary between the solar wind–dominated region and the interstellar medium. Weak cosmic rays are deflected away by the heliopause, but energetic particles penetrate the region down to the inner Solar System. (Adapted from Gosling 2007)

asteroids and Kuiper belt objects are given in Tables E.6 and E.7, and densities of some minor planets are listed in Table E.8.

The brightness of a celestial body is generally expressed as the **apparent magnitude** at visual wavelengths, $m_{\rm v}$. A 6^{th} magnitude ($m_{\rm v} = 6$) star is just visible to the naked eye in a dark sky. The magnitude scale is logarithmic (mimicking the perception of human vision), and a difference of 5 magnitudes equals a factor of 100 in brightness (i.e., a star with $m_{\rm v} = 0$ is 100 times brighter than one with $m_{\rm v} = 5$). The apparent magnitudes of planetary satellites are listed in Table E.4. Those moons with $m_{\rm v} > 20$ can only be detected with a large telescope or nearby spacecraft.

1.2.6 Heliosphere

All planetary orbits lie within the **heliosphere**, the region of space containing magnetic fields and plasma of solar origin. Figure 1.5 diagrams key components of the heliosphere. The **solar wind** consists of **plasma** (ionized gas) traveling outward from the Sun at supersonic speeds. The solar wind merges with the **interstellar medium** at the **heliopause**, the boundary of the heliosphere.

The composition of the heliosphere is dominated by solar wind protons and electrons, with a typical density of 5×10^6 protons m^{-3} at 1 AU from the Sun, decreasing as the reciprocal distance squared. These particles move outwards at speeds of $\sim$400 km s^{-1} near the solar equator but $\sim$700–800 km s^{-1} closer to the solar poles. In contrast, the local interstellar medium, at a density of less than 1×10^5 atoms m^{-3}, contains mainly hydrogen and helium atoms. The Sun's motion relative to the mean motion of neighboring stars is roughly 18 km s^{-1}. Hence, the heliosphere moves through the interstellar medium at about this speed. The heliosphere is thought to be shaped like a teardrop, with a tail in the downwind direction (Fig. 1.5). Interstellar ions and electrons generally flow around the heliosphere because they cannot cross the solar magnetic fieldlines. Neutrals, however, can enter the heliosphere, and as a result interstellar H and He atoms move through the Solar System in the downstream direction with a typical speed of $\sim$15–20 km s^{-1}.

Just interior to the heliopause is the **termination shock**, where the solar wind is slowed down. Because of variations in solar wind pressure, the location of this shock moves radially with respect to the Sun in accordance with the 11-year solar activity cycle. The *Voyager 1* spacecraft crossed the termination shock in December 2004 at a heliocentric distance of 94.0 AU; *Voyager 2* crossed the shock (multiple times) in August 2007 at $\sim$83.7 AU. The spacecraft are now in the **heliosheath**, between the termination shock and the heliopause. They are expected to reach the heliopause around the year 2015.

1.3 What Is a Planet?

The ancient Greeks referred to all moving objects in the sky as planets. To them, there were seven such objects, the Sun, the Moon, Mercury, Venus, Mars, Jupiter and Saturn. The Copernican revolution removed the Sun and Moon from the planet club, but added the Earth. Uranus and Neptune were added as soon as they were discovered in the eighteenth and nineteenth centuries, respectively.

Pluto, by far the brightest Kuiper belt object (KBO) and the first that was discovered, was officially classified as a planet from its discovery in 1930 until 2006; 1 Ceres, the first detected (in 1801) and by far the largest member of the asteroid belt, was also once considered to be a planet, as were the next few asteroids that were discovered. With the detection of other KBOs, debates began with regard to the classification of Pluto as a planet, culminating in August 2006 with the resolution by the International Astronomical Union (IAU):

- A **planet** is a celestial body that (1) is in orbit around the Sun, (2) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, and (3) has cleared the neighborhood around its orbit.
- A **dwarf planet** is a celestial body that (1) is in orbit around the Sun, (2) has sufficient mass for its self-gravity to overcome rigid body forces so that it assumes a hydrostatic equilibrium (nearly round) shape, (3) has not cleared the neighbourhood around its orbit, and (4) is not a satellite.

Just as the discoveries of small bodies orbiting the Sun have forced astronomers to decide how small an object can be and still be worthy of being classified as a planet, detections of substellar objects orbiting other stars have raised the question of an upper size limit to planethood. We adopt the following definitions, which are consistent with current IAU nomenclature:

- **Star**: self-sustaining fusion is sufficient for thermal pressure to balance gravity ($\gtrsim 0.075\ M_\odot \approx 80\ M_{♃}$ for solar composition; the minimum mass for an object to be a star is often referred to as the **hydrogen burning limit**)

- **Stellar remnant:** dead star – no more fusion (or so little that the object is no longer supported primarily by thermal pressure)
- **Brown dwarf:** substellar object with substantial deuterium fusion – more than half of the object's original inventory of deuterium is ultimately destroyed by fusion
- **Planet**: negligible fusion ($\lesssim 0.012\,M_\odot \approx 13\,M_{♃}$, with the precise value again depending on initial composition), plus it orbits one or more stars and/or stellar remnants.

1.4 Planetary Properties

All of our knowledge regarding specific characteristics of Solar System objects, including planets, moons, comets, asteroids, rings and interplanetary dust, is ultimately derived from observations, either astronomical measurements from the ground or Earth-orbiting satellites, or from close-up (often *in situ*) measurements obtained by interplanetary spacecraft. One can determine the following quantities more or less directly from observations:

(1) Orbit
(2) Mass, distribution of mass
(3) Size
(4) Rotation rate and direction
(5) Shape
(6) Temperature
(7) Magnetic field
(8) Surface composition
(9) Surface structure
(10) Atmospheric structure and composition

With the help of various theories, these observations can be used to constrain planetary properties such as bulk composition and interior structure, two attributes that are crucial elements in modeling the formation of the Solar System.

1.4.1 Orbit

In the early part of the seventeenth century, Johannes Kepler deduced three 'laws' of planetary motion directly from observations:

(1) All planets move along elliptical paths with the Sun at one focus.
(2) A line segment connecting any given planet and the Sun sweeps out area at a constant rate.
(3) The square of a planet's orbital period about the Sun, P_{orb}, is proportional to the cube of its semimajor axis, a, i.e., $P_{orb}^2 \propto a^3$.

A Keplerian orbit is uniquely specified by six orbital elements: a (semimajor axis), e (eccentricity), i (inclination), ω (argument of periapse; or ϖ for the longitude of periapse), Ω (longitude of ascending node), and f (true anomaly). These orbital elements are defined graphically in Figure 2.1 and discussed in more detail in §2.1. The first few of these elements are more fundamental than the last: a and e fully define the size and shape of the orbit, i gives the tilt of the orbital plane to some reference plane, the longitudes ϖ and Ω determine the orientation of the orbit, and f (or, indirectly, t_ϖ, the time of periapse passage) tells where the planet is along its orbit at a given time. Alternative sets of orbital elements are also possible; for instance, an orbit is fully specified by the planet's location and velocity relative to the Sun at a given time (again, six independent scalar quantities), provided the masses of the Sun and planet are known.

Kepler's laws (or more accurate versions thereof) can be derived from Newton's laws of motion and of gravity, which were formulated later in the seventeenth century (§2.1). Relativistic effects also affect planetary orbits, but they are small compared with the gravitational perturbations that the planets exert on one other (Problem 2-5).

All planets and asteroids revolve around the Sun in the direction of solar rotation. Their orbital planes generally lie within a few degrees of each other and close to the solar equator. For observational convenience, inclinations are usually measured relative to the Earth's orbital plane, which is known as the **ecliptic plane**. The Sun's equatorial plane is inclined by 7° with respect to the ecliptic plane. Among the eight major planets, Mercury's orbit is the most tilted, with $i = 7^\circ$. (However, because inclination is effectively a vector, the similarity of these two inclinations does not imply that Mercury's orbit lies within the plane of the solar equator. Indeed, Mercury's orbit is inclined by 3.4° relative to the Sun's equatorial plane.) Similarly, most major satellites orbit their planet close to its equatorial plane. Many smaller objects that orbit the Sun and the planets have much larger orbital inclinations. In addition, some comets, minor satellites and Neptune's large moon Triton orbit the Sun or planet in a **retrograde** sense (opposite to the Sun's or planet's rotation). The observed 'flatness' of most of the planetary system is explained by planetary formation models that hypothesize that the planets grew within a disk that was in orbit around the Sun (see Chapter 15).

1.4.2 Mass

The mass of an object can be deduced from the gravitational force that it exerts on other bodies.

- Orbits of moons: The orbital periods of natural satellites, together with Newton's generalization of Kepler's third law (eq. 2.18), can be used to solve for mass. The result is actually the sum of the mass of the planet and moon (plus, to a good approximation, the masses of moons on orbits interior to the one being considered), but except for the Earth/Moon and various minor planets, including Pluto/Charon, the secondaries' masses are very small compared with that of the primary. The major source of uncertainty in this method results from measurement errors in the semimajor axis; timing errors are negligible.
- What about planets without moons? The gravity of each planet perturbs the orbits of all other planets. Because of the large distances involved, the forces are much smaller, so the accuracy of this method is not high. Note, however, that Neptune was discovered as a result of the perturbations that it forced on the orbit of Uranus. This technique is still used to provide the best (albeit in some cases quite crude) estimates of the masses of some large asteroids. The perturbation method can actually be divided into two categories: short-term and long-term perturbations. The extreme example of short-term perturbations includes single close encounters between asteroids. Trajectories can be computed for a variety of assumed masses of the body under consideration and fit to the observed path of the other body. Long-term perturbations are best exemplified by masses derived from periodic variations in the relative positions of moons locked in stable orbital resonances (§2.3.2).
- Spacecraft tracking data provide the best means of determining masses of planets and moons visited because the Doppler shift and periodicity of the transmitted radio signal can be measured very precisely. The long time baselines afforded by **orbiter** missions allow much higher accuracy than **flyby** missions. The best estimates for the masses of some of the outer planet moons are those obtained by combining accurate short-term perturbation measurements from *Voyager* images with *Voyager* tracking data and/or resonance constraints from long timeline ground-based observations.
- The best estimates of the masses of some of Saturn's small inner moons were derived from the amplitude of spiral density waves they resonantly excite in Saturn's rings or of density wakes that they produce in nearby ring material. These processes are discussed in §13.4.
- Crude estimates of the masses of some comets have been made by estimating nongravitational

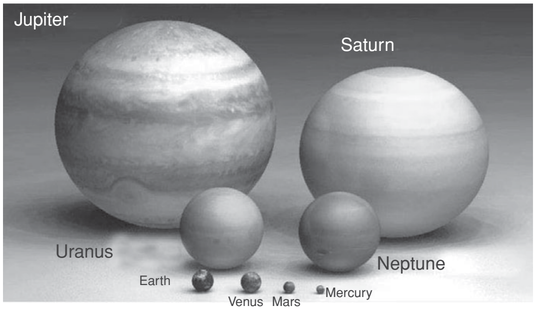


Figure 1.6 COLOR PLATE Simulated 3-D renderings of the eight planets within our Solar System. (www.lesud.com © 2011)

forces, which result from the asymmetric escape of released gases and dust (§12.2.4), and comparing them with observed orbital changes.

The gravity field of a mass distribution that is not spherically symmetric differs from that of a point source of identical mass. Such deviations, combined with the knowledge of the rotation period, can be used to estimate the degree of central concentration of mass in rotating bodies (§6.2.2). The deviation of the gravity field of an asymmetric body from that of a point mass is most pronounced, and thus most easily measured, closest to the body (§2.6). To determine the precise gravity field, one can make use of both spacecraft tracking data and the orbits of moons and/or eccentric rings.

1.4.3 Size

Bodies in the Solar System exhibit a wide range of sizes and shapes. Figure 1.6 illustrates the vast dynamic range of just the bodies considered to be planets. The size of an object can be measured in various ways:

- The diameter of a body is the product of its angular size (measured in radians) and its distance from the observer. Solar System distances are simple to estimate from orbits; however, limited resolution from Earth results in large uncertainties in angular size. Thus, other techniques often give the best results for bodies that have not been imaged at close distances by interplanetary spacecraft.
- The diameter of a Solar System body can be deduced by observing a star as it is occulted by the body. The angular velocity of the star relative to the occulting body can be calculated from orbital data, including the effects of the Earth's orbit and rotation. Multiplying the duration of an occultation as viewed from a particular observing site by both its angular velocity and its distance gives the length of a chord of the body's projected silhouette. Three well-separated chords suffice for a spherical planet. Many chords are needed if the body is irregular in shape, and observations of the same event from many widely spaced telescopes are necessary. This technique is particularly useful for small bodies that have not been visited by spacecraft. Occultations of sufficiently bright stars are infrequent and require appropriate predictions as well as significant observing campaigns in order to obtain enough chords.
- Radar echoes can be used to determine radii and shapes. The radar signal strength drops as $1/r^4$ ($1/r^2$ going to the object and $1/r^2$ returning to the antenna), so only relatively nearby objects may be studied with radar. Radar is especially useful for studying solid planets, asteroids and cometary nuclei.
- An excellent way to measure the radius of an object is to send a lander and triangulate using it together with an orbiter. This method, as well as the radar technique, also works well for terrestrial planets and satellites with substantial atmospheres.
- The size and the albedo of a body can be estimated by combining photometric observations at visible and infrared (IR) wavelengths. At visible wavelengths, one measures the sunlight reflected off the object, but at infrared wavelengths, one observes the thermal radiation from

the body itself (see Chapter 4 for a detailed discussion).

The mean density of an object can be trivially determined after its mass and size are known. The density of an object gives a rough idea of its composition, although compression at the high pressures that occur in planets and large moons must be taken into account, and the possibility of significant void space should be considered for small bodies. The low density ($\sim$1000 kg m^{-3}) of the four giant planets, for example, implies material with low mean molecular weight. Terrestrial planet densities of 3500–5500 kg m^{-3} imply rocky material, including some metal. Most of the medium and large satellites around the giant planets have densities between 1000 and 2000 kg m^{-3}, suggesting a combination of ices and rock. Comets have densities of roughly 1000 kg m^{-3} or less, indicative of rather loosely packed dirty ices.

In addition to the density, one can also calculate the escape velocity using the mass and size of the object (eq. 2.24). The escape velocity, together with temperature, can be used to estimate the ability of the planetary body to retain an atmosphere.

1.4.4 Rotation

Simple rotation is a vector quantity, related to spin angular momentum. The **obliquity** (or **axial tilt**) of a planetary body is the angle between its spin angular momentum and its orbital angular momentum. Bodies with obliquity $<90^\circ$ are said to have **prograde** rotation, and planets with obliquity $>90^\circ$ have **retrograde** rotation. The rotation of an object can be determined using various techniques:

- The most straightforward way to determine a planetary body's rotation axis and period is to observe how markings on the surface move around with the disk. Unfortunately, not all planets have such features; moreover, if atmospheric features are used, winds may cause the deduced period to vary with latitude, altitude and time.
- Planets with sufficient magnetic fields trap charged particles within their magnetospheres. These charged particles are accelerated by electromagnetic forces and emit radio waves. Because magnetic fields are not uniform in longitude and because they rotate with (presumably the bulk of) the planet, these radio signals have a periodicity equal to the planet's rotation period. For planets without detectable solid surfaces, the magnetic field period is viewed as more fundamental than the periods of cloud features (see, however, §7.3.4).
- The rotation period of a body can often be determined by periodicities observed in its **lightcurve**, which gives the total disk brightness as a function of time. Lightcurve variations can be the result of differences in albedo or, for irregularly shaped bodies, in projected area. Whereas irregularly shaped bodies produce lightcurves with two very similar maxima and two very similar minima per revolution, albedo variations have no such preferred symmetry. Thus, ambiguities of a factor of two sometimes exist in spin periods determined by lightcurve analysis. Most asteroids have double-peaked lightcurves, indicating that the major variations are due to shape, but the peaks are distinguishable from each other because of minor variations in hemispheric albedo and local topography.
- The measured Doppler shift across the disk can give a rotation period and a crude estimate of the rotation axis, provided the body's radius is known. This can be done passively in visible light or actively using radar.

The rotation periods of most objects orbiting the Sun are of the order of three hours to a few days. Mercury and Venus, both of whose rotations have almost certainly been slowed by solar tides, form exceptions with periods of 59 and 243 days, respectively. Six of the eight planets rotate in a prograde sense with obliquities of 30° or less. Venus rotates in a retrograde direction with an obliquity of 177°, and the rotation axis of Uranus is so tilted that it lies close to this planet's orbital plane. Most planetary satellites rotate synchronously with their orbital periods as a result of planet-induced tides (§2.7.2).

1.4.5 Shape

Figure 1.7a shows a close-up image of the jagged small martian moon Phobos in silhouette against the smooth limb of Mars. Many different forces together determine the shape of a body. Self-gravity tends to produce bodies of spherical shape, a minimum for gravitational potential energy. Material strength maintains shape irregularities, which may be produced by accretion, impacts or internal geological processes. Because self-gravity increases with the size of an object, larger bodies tend to be rounder. Typically, bodies with mean radii larger than ~200 km are fairly round. Smaller objects may be quite oddly shaped.

There is a relationship between a planet's rotation and its oblateness because the rotation introduces a centrifugal pseudo-force, which causes a planet to bulge out at the equator and to flatten at the poles. A perfectly fluid planet would be shaped as an oblate spheroid. Polar flattening is greatest for planets that have a low density and rapid rotation. In the case of Saturn, the flattening parameter, $\epsilon \equiv (R_e - R_p)/R_e$, where R_e and R_p are the equatorial and polar radii, respectively, is ~0.1, and polar flattening is easily discernible on some images of the planet, such as that shown in Figure 1.7b.

The shape of an object can be determined from:

- Direct imaging, either from the ground or spacecraft
- Length of chords observed by stellar occultation experiments at various sites (see §1.4.3)
- Analysis of radar echoes
- Analysis of lightcurves. Several lightcurves obtained from different viewing angles are required for accurate measurements.

(a)

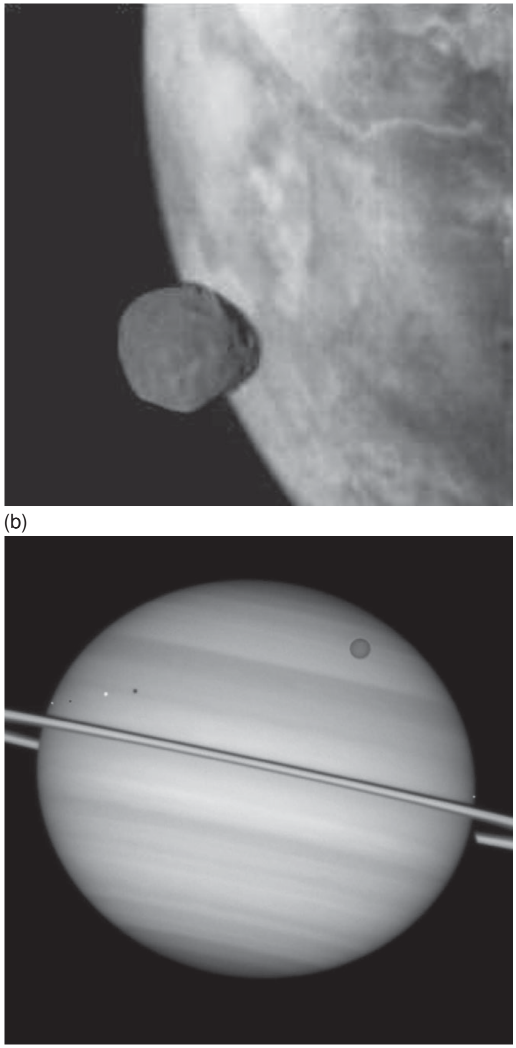


Figure 1.7 (a) Image of the small irregularly shaped moon Phobos against the background of the limb of the nearly spherical planet Mars. Phobos appears much larger relative to Mars than it actually is because the Soviet spacecraft *Phobos 2* was much closer to the moon than to the planet when it took this image. (b) *Hubble Space Telescope* image of Saturn taken on 24 February 2009 less than five months before saturnian equinox passage. The rings are seen at a low tilt angle, with the ring shadow appearing across the planet just above the rings. Four moons are seen to be transiting (partially eclipsing the planet); from left to right, they are Enceladus, Dione, Titan and Mimas; the shadows of Enceladus and Dione can also be seen. Note the pronounced oblateness of this low-density, rapidly rotating planet. (NASA/STSci/Hubble Heritage)

- The shape of the **central flash**, which is observed when the center of a body with an atmosphere passes in front of an occulted star. The central flash results from the focusing of light rays refracted by the atmosphere and can be seen only under fortuitous observing circumstances.

1.4.6 Temperature

The equilibrium temperature of a planet can be calculated from the energy balance between solar insolation and reradiation outward (see Chapter 4). However, internal heat sources provide a significant contribution to the energy balance of many planets. Moreover, there may be diurnal, latitudinal and seasonal variations in the temperature. The **greenhouse effect**, a thermal 'blanket' caused by an atmosphere that is more transparent to visible radiation (the Sun's primary output) than to infrared radiation from the planet, raises the surface temperature on some planets far above the equilibrium blackbody value. For example, because of the high albedo of its clouds, Venus actually absorbs less solar energy per unit area than does Earth; thus (as internal heat sources on these two planets are negligible compared with solar heating), the effective radiating temperature of Venus is lower than that of Earth. Nonetheless, as a consequence of the greenhouse effect, Venus's surface temperature is raised up to $\sim$730 K, well above the surface temperature on Earth.

Direct *in situ* measurements with a thermometer can provide an accurate estimate of the temperature of the accessible (outer) parts of a body. The thermal infrared spectrum of a body's emitted radiation is also a good indicator of the temperature of its surface or cloud tops. Most solid and liquid planetary material can be characterized as a nearly perfect blackbody radiator with its emission peak at near- to mid-infrared wavelengths. Analysis of emitted radiation sometimes gives different temperatures at differing wavelengths. This could be attributable to a combination of temperatures from different locations on the surface, such as pole-to-equator differences, albedo variations, or volcanic hot spots such as those seen on Io (§10.2.1). Also, the opacity of an atmosphere varies with wavelength, which allows us to remotely probe different altitudes in a planetary atmosphere.

1.4.7 Magnetic Field

Magnetic fields are created by moving charges. Currents moving through a solid medium decay quickly (unless the medium is a superconductor, which is unreasonable to expect at the high temperatures found in planetary interiors). Thus, internally generated planetary magnetic fields must either be produced by a (poorly understood) **dynamo** process, which can only operate in a fluid region of a planet (§7.4.2) or be caused by **remanent ferromagnetism**, which is a result of charges that are bound to atoms of a solid locked in an aligned configuration. Remanent ferromagnetism is not viewed to be a likely cause of large fields because, in addition to the fact that it is expected to decay away on timescales short compared with the age of the Solar System, it would require the planet to have been subjected to a nearly constant (in direction) magnetic field during the long period in which the bulk of its iron cooled through its **Curie point**. (At temperatures below the Curie point of a ferromagnetic material, the magnetic moments are partially aligned within magnet domains.) Magnetic fields may also be induced through the interaction between the solar wind (which is composed predominantly of charged particles) and conducting regions within the planet or its ionosphere.

A magnetic field may be detected directly using an *in situ* magnetometer or indirectly via radiation (radio emissions) produced by accelerating charges. The presence of localized **aurorae**, luminous disturbances caused by charged particle precipitation in a planet's upper atmosphere, is also indicative of a magnetic field. The magnetic fields of the planets can be approximated by dipoles, with perturbations to account for their irregularities. All four giant planets, as well as Earth, Mercury and Jupiter's moon Ganymede, have magnetic fields generated in their interiors. Venus and comets have magnetic fields induced by the interaction between the solar wind and charged particles in their atmosphere/ionosphere, whereas Mars and the Moon have localized crustal magnetic fields. Perturbations in Jupiter's magnetic field near Europa and Callisto are indicative of salty oceans in the interiors of these moons (§10.2). Geyser activity on Enceladus perturbs Saturn's magnetic field (§10.3.3).

1.4.8 Surface Composition

The composition of a body's surface can be derived from:

- Spectral reflectance data. Such spectra may be observed from Earth; however, spectra at ultraviolet wavelengths can only be obtained above the Earth's atmosphere.
- Thermal infrared spectra and thermal radio data. Although difficult to interpret, these measurements contain information about a body's composition.
- Radar reflectivity. Such observations can be carried out from Earth or from spacecraft that are near the body.
- X-ray and γ-ray fluorescence. These measurements may be conducted from a spacecraft in orbit around the planet (or, in theory, even a flyby spacecraft) if the body lacks a substantial atmosphere. Detailed measurements require landing a probe on the body's surface.
- Chemical analysis of surface samples. This can be performed on samples brought to Earth by natural processes (meteorites) or spacecraft, or (in less detail) by *in situ* analysis using spacecraft. Other forms of *in situ* analysis include mass spectroscopy and electrical and thermal conductivity measurements.

The compositions of the planets, asteroids and satellites show a dependence on heliocentric distance, with the objects closest to the Sun having the largest concentrations of dense materials (which tend to be **refractory**, i.e., have high melting and boiling temperatures) and the smallest concentration of ices (which are much more **volatile**, i.e., have much lower melting and boiling temperatures).

1.4.9 Surface Structure

The surface structure varies greatly from one planet or moon to another. There are various ways to determine the structure of a planet's surface:

- Structure on large scales (e.g., mountains) can be detected by imaging, either passively in the visible/infrared/radio or actively using radar imaging techniques. It is best to have imaging available at more than one illumination angle in order to separate tilt-angle (slope) effects from albedo differences.
- Structure on small scales (e.g., grain size) can be deduced from the radar echo brightness and the variation of reflectivity with **phase angle**, the angle between the illuminating Sun and the observer as seen from the body. The brightness of a body with a size much larger than the wavelength of light at which it is observed generally increases slowly with decreasing phase angle. For very small phase angles, this increase can be much more rapid, a phenomenon referred to as the **opposition effect**.

1.4.10 Atmosphere

Most of the planets and some satellites are surrounded by significant atmospheres. The giant planets Jupiter, Saturn, Uranus and Neptune are basically huge fluid balls, and their atmospheres are dominated by H_2 and He. Venus has a very dense CO_2 atmosphere, with clouds so thick that one cannot see its surface at visible wavelengths;

Earth has an atmosphere consisting primarily of N_2 (78%) and O_2 (21%), and Mars has a more tenuous CO_2 atmosphere. Saturn's satellite Titan has a dense nitrogen-rich atmosphere, which is intriguing because it contains many kinds of organic molecules. Pluto and Neptune's moon Triton each have a tenuous atmosphere dominated by N_2, and the atmosphere of Jupiter's volcanically active moon Io consists primarily of SO_2. Mercury and the Moon each have an extremely tenuous atmosphere ($\lesssim 10^{-12}$ bar); Mercury's atmosphere is dominated by atomic O, Na and He, and the main constituents in the Moon's atmosphere are He and Ar. The gaseous components of cometary comae are essentially temporary atmospheres in the process of escaping.

The composition and structure (temperature–pressure profile) of an atmosphere can be determined from spectral reflectance data at visible wavelengths, thermal spectra and photometry at infrared and radio wavelengths, stellar occultation profiles, *in situ* mass spectrometers and attenuation of radio signals sent back to Earth by atmospheric/surface probes.

1.4.11 Interior

The interior of a planet is not directly accessible to observations. However, with help of the observable parameters discussed earlier, one can derive information on a planet's bulk composition and its interior structure.

The **bulk composition** is not an observable attribute, except for extremely small bodies, such as meteorites, that we can actually take apart and analyze (see Chapter 11). Thus, we must deduce bulk composition from a variety of direct and indirect clues and constraints. The most fundamental constraints are based on the mass and the size of the planet. Using only these constraints together with material properties derived from laboratory data and quantum mechanical calculations, it can be shown that Jupiter and Saturn are composed mostly of hydrogen, simply because all other elements are too dense to fit the constraints (unless the internal temperature is much higher than is consistent with the observed effective temperature in a quasi-steady state). However, this method only gives definitive results for planets composed primarily of the lightest element. For all other bodies, bulk composition is best estimated from models that include mass and radius as well as the composition of the surface and atmosphere, the body's heliocentric distance (location is useful because it gives us an idea of the temperature of the region during the planet-formation epoch and thus which elements were likely to condense), together with reasonable assumptions of cosmogonic abundances (§1.5, Table 3.1 and Chapters 11 and 15).

The **internal structure** of a planet can be derived to some extent from its gravitational field and rotation rate. From these parameters, one can estimate the degree of concentration of the mass at the planet's center. The gravitational field can be determined from spacecraft tracking and the orbits of satellites or rings. Detailed information on the internal structure of a planet with a solid surface may be obtained if seismometers can be placed on its surface, as was done for the Moon by *Apollo* astronauts. The velocities and attenuations of seismic waves propagating through the planet's interior depend on density, rigidity and other physical properties, which in turn depend on composition, as well as on pressure, temperature and time. Reflection and refraction off internal boundaries provide information on layering. The free oscillation periods of gaseous planets can, in theory, also provide clues to internal properties, just as **helioseismology**, the study of solar oscillations, now provides important information about the Sun's interior. Evidence of volcanism and plate tectonics constrain the thermal environment below the surface. Energy output provides information on the thermal structure of a planet's interior.

The response of moons that are subject to significant time-variable tidal deformations depends on their internal structure. Repeated observations of such moons can reveal internal properties, including in some cases the presence of a subterranean fluid layer. Combining altitude and gravity field measurements could give indications about lateral inhomogeneities under the surface of icy moons and thus, for instance, indicate volcanic sources and tectonic structures.

Magnetic fields are produced by moving charges. Although a small magnetic field such as the Moon's may be the result of remanent ferromagnetism, substantial planetary magnetic fields are thought to require a conducting fluid region within the planet's interior. Whereas centered dipole fields are probably produced in or near the core of the planet, highly irregular offset fields are likely to be produced closer to the planet's surface.

1.5 Formation of the Solar System

The nearly planar and almost circular orbits of the planets in our Solar System argue strongly for planetary formation within a flattened circumsolar disk. Astrophysical models suggest that such disks are a natural byproduct of star formation from the collapse of rotating cores of molecular clouds. Observational evidence for the presence of disks of Solar System dimensions around young stars has increased substantially in recent years, and infrared excesses in the spectra of young stars suggest that the lifetimes of protoplanetary disks range from 10^6–10^7 years.

Our galaxy contains many molecular clouds, most of which are several orders of magnitude larger than our Solar System. **Molecular clouds** are the coldest and densest parts of the interstellar medium. They are inhomogeneous, and the densest parts of molecular clouds are referred to as **cores**. These are the sites in which star formation occurs at the current epoch. Even a very slowly rotating molecular cloud core has far too much spin angular momentum to collapse down to an object of stellar dimensions, so a significant fraction of the material in a collapsing core falls onto a rotationally supported disk orbiting the pressure-supported (proto)star. Such a disk has the same initial elemental composition as the growing star. At sufficient distances from the central star, it is cool enough for $\sim$1%–2% of this material to be in solid form, either remnant interstellar grains or condensates formed within the disk. This dust is primarily composed of rock-forming compounds within a few AU of a 1 $M_\odot$ star, but in the cooler, more distant regions, the amount of ices (e.g., H_2O, CH_4, CO) present in solid form is comparable to that of rocky solids.

During the infall stage, the disk is very active and probably highly turbulent as a result of the mismatch of the specific angular momentum of the gas hitting the disk with that required to maintain Keplerian rotation. Gravitational instabilities and viscous and magnetic forces may add to this activity. When the infall slows substantially or stops, the disk becomes more quiescent. Interactions with the gaseous component of the disk affect the dynamics of small solid bodies, and the growth from micrometer-sized dust to kilometer-sized planetesimals remains poorly understood. Meteorites (see Chapter 11), minor planets and comets (see Chapter 12), most of which were never incorporated into bodies of planetary dimensions, best preserve a record of this important period in Solar System development.

The dynamics of larger solid bodies within protoplanetary disks are better characterized. The primary perturbations on the Keplerian orbits of kilometer-sized and larger planetesimals in protoplanetary disks are mutual gravitational interactions and physical collisions. These interactions lead to accretion (and in some cases erosion and fragmentation) of planetesimals. Eventually, solid bodies agglomerated into the terrestrial planets in the inner Solar System and into planetary cores several times the mass of the Earth in the outer Solar System. These massive cores were able to gravitationally attract and retain substantial amounts of gaseous material from the solar nebula. In contrast, terrestrial planets were not massive enough to attract and retain such gases, and the gases in their current thin atmospheres are derived from material that was incorporated in solid planetesimals.

The planets in our Solar System orbit close enough to one another that the final phases of planetary growth could have involved the merger or ejection of planets or planetary embryos on unstable orbits. However, the low eccentricities of the orbits of the outer planets imply that some damping process, such as accretion/ejection of numerous small planetesimals or interactions with residual gas within the protoplanetary disk, must also have been involved.

As researchers learn more about the individual bodies and classes of objects in our Solar System, and as simulations of planetary growth become more sophisticated, theories about the formation of our Solar System are being revised and (we hope) improved. The detection of planets around other stars has presented us with new challenges to develop a unified theory of planet formation that is more generally applicable. We discuss these theories in more detail in Chapter 15.

Key Concepts

- Planets are the wanderers of the night sky, changing in position relative to the 'fixed' stars.
- The study of the motions of the planets dates back thousands of years, but most of our knowledge about planets and smaller bodies within our Solar System has been obtained during the space age.
- The Sun dominates our Solar System in most respects, followed by Jupiter, then Saturn and after that the pair Uranus and Neptune.
- A wide variety of techniques are used to observe the properties of planetary bodies. However, some planetary characteristics, such as interior composition, cannot at present be directly observed and can only be deduced from theoretical modeling.
- When a molecular cloud core collapses, the inner portion becomes a star. Molecular cloud material with high angular momentum falls into a disk around that star and is available for planet formation.
- Whereas planets grow by accretion of small bodies into larger ones, stars form via the collapse of large clouds into smaller objects.

Further Reading

More extensive and technical accounts of most of the topics presented in this book (other than those connected to life) can be found in our graduate-level textbook:

de Pater, I., and J.J. Lissauer, 2010. *Planetary Sciences*, 2nd Edition. Cambridge University Press, Cambridge. 647pp.

A good nontechnical overview of our planetary system, complete with many beautiful color pictures, is given by:

Beatty, J.K., C.C. Peterson, and A. Chaikin, Eds., 1999. *The New Solar System*, 4th Edition. Sky Publishing Co., Cambridge, MA and Cambridge University Press, Cambridge. 421pp.

A terse but detailed overview, including reproductions of paintings of various Solar System objects by the authors, is provided by:

Miller, R., and W.K. Hartmann, 2005. *The Grand Tour: A Traveler's Guide to the Solar System*, 3rd Edition. Workman Publishing, New York. 208pp.

An overview of the Solar System emphasizing atmospheric and space physics is given by:

Encrenaz, T., J.-P. Bibring, M. Blanc, M.-A. Barucci, F. Roques, and Ph. Zarka, 2004. *The Solar System*, 3rd Edition. Springer-Verlag, Berlin. 512pp.

Two good overview texts aimed at college students not majoring in science are:

Morrison, D., and T. Owen, 2003. *The Planetary System*, 3rd Edition. Addison-Wesley Publishing Company, New York. 531pp.
Hartmann, W.K., 2005. *Moons and Planets*, 5th Edition. Brooks/Cole, Thomson Learning, Belmont, CA. 428pp.

Short summaries of a multitude of topics, ranging from mineralogy to black holes, at a level of sophistication a bit higher than that of this book, are provided by:

Cole, G.H.A., and M.M. Woolfson, 2002. *Planetary Science: The Science of Planets Around Stars*, Institute of Physics Publishing, Bristol and Philadelphia. 508pp.

Chemical processes on planets and during planetary formation are covered in some detail by:

Lewis, J.S., 2004. *Physics and Chemistry of the Solar System*, 2nd Edition. Elsevier, Academic Press, San Diego. 684pp.

The following encyclopedia forms a nice complement to this book:

McFadden, L., P. R. Weissman, and T.V. Johnson, Eds., 2007. *Encyclopedia of the Solar System*, 2nd Edition. Academic Press, San Diego. 982pp.

Extensive planetary data tables can be found in:

Yoder, C.F., 1995. Astrometric and geodetic properties of Earth and the Solar System. In *Global Earth Physics: A Handbook of Physical Constants*. AGU Reference Shelf 1, American Geophysical Union, 1–31.

For updated information, see http://ssd.jpl.nasa.gov.

A collection of beautiful images of planets and astrobiology can be found at http://fettss.arc.nasa.gov/collection/.

Problems

1-1. Because the distances between the planets are much larger than planetary sizes, very few diagrams or models of the Solar System are completely to scale. However, imagine that you are asked to give an astronomy lecture and demonstration to your niece's second-grade class, and you decide to illustrate the vastness and near emptiness of space by constructing a scale model of the Solar System using ordinary objects. You begin by selecting a (1-cm-diameter) marble to represent the Earth.

(a) What other objects can you use, and how far apart must you space them?

(b) Proxima Centauri, the nearest star to the Solar System, is 4.2 light years distant; where, in your model, would you place it?

1-2. The satellite systems of the giant planets are often referred to as 'miniature solar systems'. In this problem, you will make some calculations comparing the satellite systems of Jupiter, Saturn and Uranus with the Solar System.

(a) Calculate the ratio of the sum of the masses of the planets with that of the Sun and similar ratios for the jovian, saturnian and uranian systems using the respective planet as the primary mass.

(b) Calculate the ratio of the sum of the orbital angular momenta of the planets to the rotational angular momentum of the Sun. You can assume circular orbits at zero inclination for all planets and ignore the effects of planetary rotation and the presence of satellites. The Sun rotates differentially, with a mean rotation period of 25.4 days.

(c) Repeat the calculation in (b) for the jovian, saturnian and uranian systems using the respective planet as the primary mass.

(d) Calculate the orbital semimajor axes of the planets in terms of solar radii and the orbital semimajor axes of Jupiter's moons in jovian radii. How would a scale model of the jovian system compare with the model of the Solar System in Problem 1-1?

1-3. **(a)** Standing on the surface or floating in the atmosphere of which Solar System body would you see the brightest object in the nighttime sky? Justify your answer.

(b) Same question but assume that you are standing on a body with a solid surface and a significant atmosphere.

1-4. A planet that keeps the same hemisphere pointed towards the Sun must rotate once per orbit in the prograde direction.

(a) Draw a diagram to demonstrate this fact. Whereas the rotation period (in an inertial frame) or **sidereal day** for such a planet is equal to its orbital period, the length of a **solar day** on such a planet is infinite.

(b) Earth rotates in the prograde direction. How many times must Earth rotate per orbit for there to be 365.24 solar days per year? Verify your result by comparing the length of Earth's sidereal rotation period (Table E.2) with the length of a mean solar day.

(c) If a planet rotated once per orbit in the retrograde direction, how many solar days would it have per orbit?

(d)* Determine a general formula relating the lengths of solar and sidereal days on a planet. Use your formula to compute the lengths of solar days on Mercury, Venus, Mars and Jupiter.

(e)* For a planet on an eccentric orbit, the length of either the solar day or the sidereal day varies on an annual cycle. Which one varies, and why? Calculate the length of the longest such day on Earth. This longest day is how much longer than the mean day of its type?

(Note: The Earth's obliquity causes variations in the rate of apparent motion of the Sun along the equator, which also produce variations in the length of the day. The **equation of time** accounts for both types of variations and enables accurate calculation of the time using a sundial.)

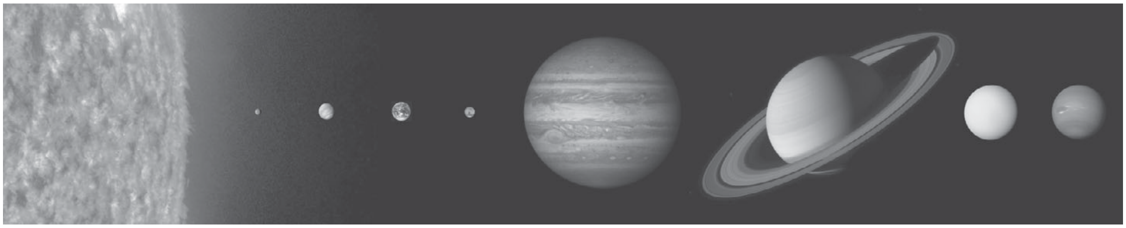
1-5. A **total solar eclipse** occurs when the Moon blocks the entire disk of the Sun, allowing the observer to view only the Sun's extended atmosphere, the **corona**. An **annular eclipse** occurs when the Moon obscures the central portion of the Sun but a narrow annulus of the Sun's photosphere can be seen surrounding the Moon.

(a) Using the data in Tables E.4, E.5 and C.5, show that the eccentricities of the orbits of Earth about the Sun and the Moon about the Earth make it possible for both types of eclipse to be viewed from the surface of Earth.

(b) Which occur more frequently, total solar eclipses or annular eclipses? Why?

CHAPTER 2

Dynamics



The Planets move one and the same way in Orbs concentric, some inconsiderable Irregularities excepted, which may have arisen from the mutual Actions of Planets upon one another, and which will be apt to increase, till this System wants a Reformation.

Isaac Newton, *Opticks*

Dynamical studies of planetary bodies characterize their motions, including rotation and deformation of bodies resulting from tidal distortions. Dynamics is the oldest of the planetary sciences. Gravitational interactions determine how the distance of a planet from the Sun varies with time and thus how much solar radiation the planet intercepts. Rotation rates determine the length of the day; obliquity influences pole-equator temperature differences and seasonal variations. Tidal heating produces extensive volcanism on bodies such as Jupiter's moon Io (see Fig. 10.4).

The history of observational studies of, and kinematical models for, planetary motions dates back to antiquity. Modern planetary dynamics began in the seventeenth century. In the first decades of that century, Johannes Kepler conducted an extensive analysis of planetary observations that had been made in the previous decades by Tycho Brahe. Towards the end of the seventeenth century, Isaac Newton provided a firm basis for dynamical studies by discovering physical laws that govern the motions of objects on Earth as well as in the heavens. Albert Einstein's (twentieth century) theory of relativity fundamentally modified the underlying theories of motion and gravity, but the magnitude of relativistic corrections to planetary motions is generally quite small (Problems 2-4 and 2-5).

In 1687, Newton showed that the relative motion of two spherically symmetric bodies resulting from their mutual gravitational attraction is described by simple conic sections: ellipses for bound orbits and parabolas and hyperbolas for unbound trajectories. However, the introduction of additional gravitating bodies produces a rich variety of dynamical phenomena even though the basic interactions between pairs of objects can be straightforwardly described.

In this chapter, we describe the basic orbital properties of Solar System objects (planets, moons, minor bodies and dust) and their mutual interactions. We also provide several examples of important dynamical processes that occur in the Solar System and lay the groundwork for describing some of the phenomena that are considered in other chapters of this book.

We begin in §2.1 with an overview of the **two-body problem**, i.e., the relative motion of an isolated pair of spherically symmetric objects that are gravitationally attracted to one another. Our discussion introduces Kepler's laws, Newton's laws and the terminology used to describe planetary orbits. In the next three sections, we discuss the consequences of gravitational interactions among larger numbers of bodies. We consider the dynamics of spherically symmetric objects of finite size in §2.5. We relax the assumption of spherical symmetry in §2.6 to analyze the dynamics of rotating planets and of orbits about them, and we consider the effects of tidal forces on deformable bodies in §2.7. Although gravity is the dominant force on the motions of large bodies in the Solar System, electromagnetic forces such as radiation pressure substantially affect the motions of small objects, which have larger surface area to mass ratios than do large objects; we discuss such forces in §2.8. We conclude the chapter with a brief overview of orbits about a mass-losing star, which may be important for very young and very old planetary systems.

2.1 The Two-Body Problem

All bodies in the Universe are subject to the gravitational attraction of all other bodies. But for many planetary science applications, the trajectory of one body is well approximated by considering just the gravitational force exerted on it by a single other body. We describe the analysis of this elementary yet nontrivial problem and various applications in this section.

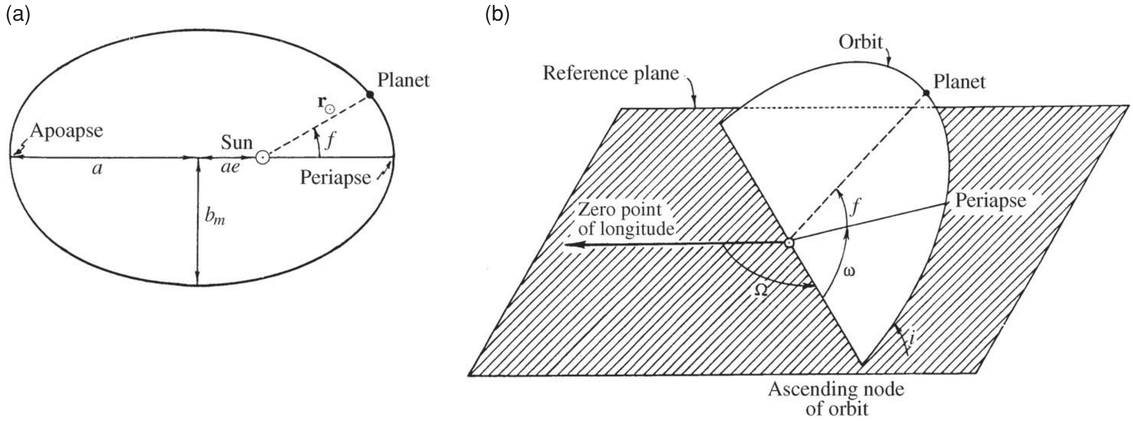


Figure 2.1 (a) Geometry of an elliptical orbit. The Sun is at one focus, and the vector $\mathbf{r}_\odot$ denotes the instantaneous heliocentric location of the planet (i.e., $r_\odot$ is the planet's distance from the Sun). The semimajor axis of the ellipse is a, e denotes its eccentricity, and b_m is the ellipse's semiminor axis. The true anomaly, f, is the angle between the planet's perihelion and its instantaneous position. (b) Geometry of an orbit in three dimensions; i is the inclination of the orbit, Ω is the longitude of the ascending node and ω is the argument of periapse. (Adapted from Hamilton 1993)

2.1.1 Kepler's Laws of Planetary Motion

By careful analysis of the observed orbits of the planets, Kepler deduced his three 'laws' of planetary motion:

(1) All planets move along elliptical paths with the Sun at one focus. We can express the **heliocentric distance**, $r_\odot$ (i.e., the planet's distance from the Sun), as

$$r_\odot = \frac{a(1-e^2)}{1+e\cos f}, \tag{2.1}$$

with a the **semimajor axis** (average of the minimum and maximum heliocentric distances). The **eccentricity** of the orbit, $e \equiv (1 - b_m^2/a^2)^{1/2}$, where $2b_m$ is the minor axis of the ellipse. The **true anomaly**, f, is the angle between the planet's **perihelion** (where it is closest to the Sun) and its instantaneous position. These quantities are displayed graphically in Figure 2.1a.

(2) A line connecting any given planet and the Sun sweeps out area, $\mathcal{A}$, at a constant rate:

$$\frac{d\mathcal{A}}{dt} = \text{constant}. \tag{2.2}$$

The value of this constant rate differs from one planet to the next. Kepler's second law is illustrated in Figure 2.2.

(3) The square of a planet's orbital period about the Sun (in years), P_{yr}, is equal to the cube of its semimajor axis (in AU), a_{AU}:

$$P_{yr}^2 = a_{AU}^3. \tag{2.3}$$

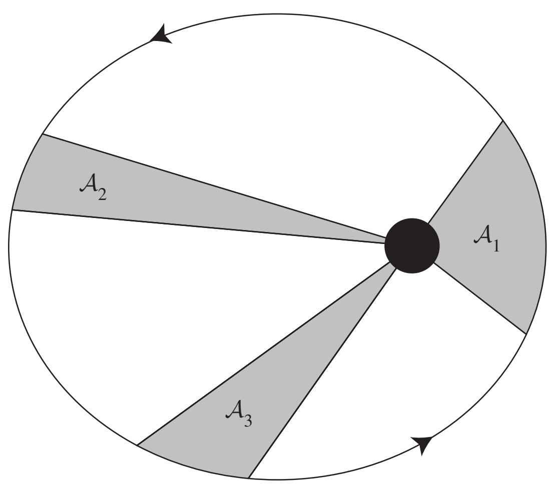


Figure 2.2 Schematic illustration of Kepler's second law. (Murray and Dermott 1999)

2.1.2 Newton's Laws of Motion and Gravity

Isaac Newton developed the first physical model that explained the motion of objects on Earth and in the heavens using a single, unified theory. Newton's theory includes four 'laws', three explaining motion and the fourth quantifying the gravitational force.

Newton's first law concerns inertia: A body remains at rest or in uniform motion unless a force is exerted upon it.

Consider a body of mass m_1 at instantaneous location $\mathbf{r}_1$ with instantaneous velocity $\mathbf{v}_1 \equiv d\mathbf{r}_1/dt$ and hence momentum $m_1\mathbf{v}_1$. The acceleration produced by a net force $\mathbf{F}_1$ is given by Newton's second law of motion:

$$\frac{d(m_1\mathbf{v}_1)}{dt} = \mathbf{F}_1. \tag{2.4}$$

Newton's third law states that for every action there is an equal and opposite reaction; thus, the force on each object of a pair due to the other object is equal in magnitude but opposite in direction:

$$\mathbf{F}_{12} = -\mathbf{F}_{21}, \tag{2.5}$$

where $\mathbf{F}_{ij}$ represents the force exerted by body j on body i.

Newton's universal law of gravity states that a second body of mass m_2 at position $\mathbf{r}_2$ exerts an attractive force on the first body given by

$$\mathbf{F}_{g12} = -\frac{Gm_1m_2}{r^2}\hat{\mathbf{r}}, \tag{2.6}$$

where $\mathbf{r} \equiv \mathbf{r}_1 - \mathbf{r}_2$ is the vector distance from particle 2 to particle 1, G is the gravitational constant and $\hat{\mathbf{r}} \equiv \mathbf{r}/r$.

Although Kepler's laws were originally deduced from careful observation of planetary motion, they were subsequently shown to be derivable from Newton's laws of motion together with his universal law of gravity. We present portions of this derivation (using modern mathematics and notation) below.

2.1.3 Reduction of the Two-Body Problem to the One-Body Problem

The equation for the relative motion of two mutually gravitating bodies can be derived from Newton's laws. Consider two mutually gravitating bodies of masses m_1 and m_2 and positions $\mathbf{r}_1$ and $\mathbf{r}_2$. Newton's second law of motion (eq. 2.4) can be combined with his law of gravitation (eq. 2.6) to yield the following two equations that govern the motion of these bodies:

$$m_1\frac{d^2\mathbf{r}_1}{dt^2} = -\frac{Gm_1m_2}{|\mathbf{r}_1 - \mathbf{r}_2|^3}(\mathbf{r}_1 - \mathbf{r}_2), \tag{2.7}$$

$$m_2\frac{d^2\mathbf{r}_2}{dt^2} = -\frac{Gm_1m_2}{|\mathbf{r}_2 - \mathbf{r}_1|^3}(\mathbf{r}_2 - \mathbf{r}_1). \tag{2.8}$$

To separate the motion of the center of mass from the relative motion of the two bodies, we apply the coordinate transformation $\mathbf{x} \equiv (m_1\mathbf{r}_1 + m_2\mathbf{r}_2)/(m_1 + m_2)$, $\mathbf{r} \equiv \mathbf{r}_1 - \mathbf{r}_2$. Substitution and simple algebraic manipulation yields:

$$(m_1 + m_2)\frac{d^2\mathbf{x}}{dt^2} = \mathbf{0}, \tag{2.9}$$

and

$$\frac{d^2\mathbf{r}}{dt^2} = -\frac{GM}{r^2}\hat{\mathbf{r}}, \tag{2.10}$$

where $M \equiv m_1 + m_2$. Equation (2.9) implies that the center of mass of the system does not accelerate and therefore moves at constant velocity. Equation (2.10) describes the acceleration of the relative position of the two bodies, $\mathbf{r}$.

Thus, the relative motion of the two bodies is completely equivalent to that of a particle orbiting a *fixed* central mass M. This reduces the two-body problem to an equivalent one-body problem.

2.1.4* Generalization of Kepler's Laws

Having reduced the two-body problem to an equivalent one-body problem, we proceed with the derivation of (Newton's generalization of) Kepler's laws. As $\mathbf{v} \equiv d\mathbf{r}/dt$, vector calculus manipulation implies that

$$\frac{d}{dt}(\mathbf{r} \times \mathbf{v}) = \mathbf{r} \times \frac{d\mathbf{v}}{dt} + \frac{d\mathbf{r}}{dt} \times \mathbf{v}$$
$$= \mathbf{r} \times \frac{d^2\mathbf{r}}{dt^2} + \mathbf{v} \times \mathbf{v} = \mathbf{0}, \tag{2.11}$$

where the first two equalities in equation (2.11) are valid in general and the last equality uses the force law given by equation (2.6). Equation (2.11) implies that the angular momentum, $\mathbf{L}$, which is given by:

$$\mathbf{L} \equiv \mathbf{r} \times m\mathbf{v}, \tag{2.12}$$

is conserved, i.e.,

$$\frac{d\mathbf{L}}{dt} = 0. \tag{2.13}$$

In polar coordinates, the expression for the magnitude of the angular momentum is just $L = mrv_\theta$. The rate of sweeping is

$$\frac{d\mathcal{A}}{dt} = \frac{rv_\theta}{2} = \frac{L}{2m}. \tag{2.14}$$

Conservation of angular momentum (eq. 2.12) thus yields the Newtonian generalization of Kepler's second law:

(2) A line connecting two bodies (as well as lines from each body to the center of mass) sweeps out area at a constant rate. The value of this constant is given by equation (2.14).

The derivation of the generalized versions of Kepler's first and third laws is mathematically straightforward but rather tedious. The details of these derivations are presented in many books and are available on the web. We therefore only sketch the procedure and quote the results below.

To derive Kepler's first law, take the dot product of $\mathbf{v}$ with equation (2.10) to derive the equation of conservation of energy per unit mass. Integrate your result to determine an expression for the specific energy of the system, E. Express your answer in polar coordinates and solve for dr/dt. Take the reciprocal; multiply both sides by $d\theta/dt$; and then use the magnitude of the specific angular momentum, L, to eliminate the angular velocity from your expression, yielding the following purely spatial relationship for the orbit:

$$\frac{d\theta}{dr} = \frac{1}{r}\left(\frac{2Er^2}{L^2} + \frac{2GMr}{L^2} - 1\right)^{-1/2}. \tag{2.15}$$

Integrate equation (2.15) and solve for r. Set the constant of integration equal to $-\pi/2$, define $r_0 \equiv L^2/(GM)$ and use the relationship $e = \left(1 + (2EL^2)/(G^2M^2)\right)^{1/2}$ to obtain:

$$r = \frac{r_0}{1 + e\cos\theta}. \tag{2.16}$$

For $0 \le e < 1$, equation (2.16) represents an ellipse in polar coordinates. Thus, Kepler's first law is also precise in the two-body Newtonian approximation, although the Sun itself is not fixed in space. Note that if $E = 0$, then $e = 1$ and equation (2.16) describes a parabola, and if $E > 0$, then $e > 1$ and the orbit is hyperbolic. The generalized form of Kepler's first law reads:

(1) The two bodies move along elliptical paths, with one focus of each ellipse located at the center of mass (CM) of the system,

$$\mathbf{r}_{\mathrm{CM}} = \frac{m_1\mathbf{r}_1 + m_2\mathbf{r}_2}{M}. \tag{2.17}$$

To derive Kepler's third law, begin by showing that the semimajor and semiminor axes of the ellipse given by equation (2.16) are $a = r_0/(1 - e^2)$ and $b = r_0/(1 - e^2)^{1/2}$, respectively. Determine the orbital period, P, by setting the integral of $d\mathcal{A}/dt$ equal to the area of the ellipse, πab. The resulting generalized form of Kepler's third law is:

(3) The orbital period of a pair of bodies about their mutual center of mass is given by

$$P_{\mathrm{orb}}^2 = \frac{4\pi^2 a^3}{GM}. \tag{2.18}$$

Note that the result given in equation (2.18) differs from Kepler's third law by replacing the Sun's mass, m_1, by the sum of the masses of the Sun and the planet, M.

2.1.5 Orbital Elements

The Sun contains more than 99.8% of the mass of the known Solar System. The gravitational force exerted by a body is proportional to its mass (eq. 2.6), so to an excellent first approximation we can regard the motion of the planets and many other bodies as being solely influenced by a fixed central pointlike mass. For objects such as the planets, which are bound to the Sun and hence cannot go arbitrarily far from the central mass, the general solution for the orbit is the ellipse described by equation (2.1).

The orbital plane, although fixed in space, can be arbitrarily oriented with respect to whatever reference plane we have chosen. This reference plane is usually taken to be either the Earth's orbital plane about the Sun, which is called the **ecliptic**, or the equatorial plane of the largest body in the system, or the **invariable plane** (the plane perpendicular to the total angular momentum of the system). The Solar System's invariable plane is nearly coincident with the plane of Jupiter's orbit, which is inclined by 1.3° relative to the ecliptic. In this book, we follow standard conventions and measure inclinations of heliocentric orbits with respect to the ecliptic plane and inclinations of planetocentric orbits relative to the planet's equator.

The terminology and variables used to describe orbits are shown in Figure 2.1. The **inclination**, i, of the orbit is the angle between the reference plane and the orbital plane; i can range from 0° to 180°. Conventionally, secondaries orbiting in the same direction as the primary rotates are defined to have inclinations from 0° to 90° and are said to be on **prograde** (or **direct**) orbits. Secondaries orbiting in the opposite direction are defined to have $90^\circ < i \leq 180^\circ$ and said to be on **retrograde** orbits. For heliocentric orbits, the Earth's orbital plane rather than the Sun's equator is usually taken as the reference. The intersection of the orbital and reference planes is called the **line of nodes**, and the orbit pierces the reference plane at two locations – one as the body passes upward through the plane (the **ascending node**) and one as it descends (the **descending node**). A fixed direction in the reference plane is chosen, and the angle to the direction of the orbit's ascending node is called the **longitude of the ascending node**, Ω.

The angle between the line to the ascending node and the line to the direction of **periapse** (the point on the orbit when the two bodies are closest, which is referred to as **perihelion** for orbits about the Sun and **perigee** for orbits about the Earth) is called the **argument of periapse**, ω. For heliocentric orbits, Ω and ω are measured eastward from the vernal equinox. The **vernal equinox** is the great circle through the celestial poles that crosses the equator at the location of the Sun on the first day of spring. Finally, the true anomaly, f, specifies the angle between the planet's periapse and its instantaneous position. Thus, the six **orbital elements**, a, e, i, Ω, ω and f, uniquely specify the location of the object in space (Fig. 2.1). The first three quantities, a, e, and i, are often referred to as the **principal orbital elements** because they describe the size, shape and tilt of the orbit.

For two bodies with known masses, specifying the elements of the relative orbit and the positions and velocities of the center of mass is equivalent to specifying the positions and velocities of both bodies. Alternative (sets of) orbital elements are often used for convenience. For example, the **longitude of periapse**,

$$\varpi \equiv \Omega + \omega, \tag{2.19a}$$

can be used in place of ω. The time of perihelion passage, t_ϖ, is commonly used instead of f as an alternative way by which to specify the location of the particle along its orbital path. The **mean motion** (average angular speed),

$$n \equiv \frac{2\pi}{P_{\rm orb}}, \tag{2.19b}$$

and the **mean longitude**,

$$\lambda = n(t - t_{\varpi}) + \varpi, \tag{2.19c}$$

are also used to specify orbital properties.

2.1.6 Bound and Unbound Orbits

For a pair of bodies to travel on a circular orbit about their mutual center of mass, they must be pulled towards one another enough to balance inertia. Quantitatively, gravity must balance the centrifugal pseudoforce that is present if the problem is viewed as a steady state in the frame rotating with the angular velocity of the two bodies, n. The **centripetal force** necessary to keep an object of mass m in a circular orbit of radius r with speed $v_{\rm c}$ is

$$\mathbf{F}_{\rm c} = mn^2\mathbf{r} = \frac{m{v_{\rm c}}^2}{r}\hat{\mathbf{r}}. \tag{2.20}$$

Equating this to the gravitational force exerted by the central body of mass M, we find that the speed of a circular orbit is

$$v_{\rm c} = \sqrt{\frac{GM}{r}}. \tag{2.21}$$

The total energy of the system, E, is a conserved quantity:

$$E = \frac{1}{2}mv^2 - \frac{GMm}{r} = -\frac{GMm}{2a}, \tag{2.22}$$

where the first term in the middle expression is the kinetic energy of the system and the second term is potential energy. For circular orbits, the second equality in equation (2.22) follows immediately from equation (2.21).

If $E < 0$, the absolute value of the potential energy of the system is larger than its kinetic energy, and the system is **bound**: The body orbits the central mass on an elliptical path. Simple manipulation of equation (2.22) yields an expression for the velocity along an elliptical orbit at each radius r:

$$v^2 = GM\left(\frac{2}{r} - \frac{1}{a}\right). \tag{2.23}$$

Equation (2.23) is known as the **vis viva equation**. If $E > 0$, the kinetic energy is larger than the absolute value of the potential energy, and the system is **unbound**. The orbit is then described mathematically as a hyperbola. If $E = 0$, the kinetic and potential energies are equal in magnitude, and the orbit is a parabola. By setting the total energy (eq. 2.22) equal to zero, we can calculate the **escape velocity** (alternatively referred to as the **escape speed**) at any separation:

$$v_{\rm e} = \sqrt{\frac{2GM}{r}} = \sqrt{2}\, v_{\rm c}. \tag{2.24}$$

As noted earlier, the orbit in the two-body problem is an ellipse, parabola or hyperbola corresponding to the energy being negative, zero or positive, respectively. These curves are known collectively as **conic sections** and are illustrated in Figure 2.3. The generalization of equation (2.1) to include unbound as well as bound orbits is

$$r = \frac{\zeta}{1 + e\cos f}, \tag{2.25}$$

where r and f have the same meaning as in equation (2.1), e is the **generalized eccentricity** and ζ is a constant. Bound orbits have $e < 1$ and $\zeta = a(1 - e^2)$, but the generalized eccentricity can take any non-negative value. For elliptical orbits, the generalized eccentricity is no different from the eccentricity defined in §2.1.1. For a parabola, $e = 1$ and $\zeta = 2q$, where q is the **pericentric separation**, i.e., the distance of closest approach. For a hyperbola, $e > 1$ and $\zeta = q(1 + e)$; $e \gg 1$ signifies a hyperbola with only a slight bend, nearly

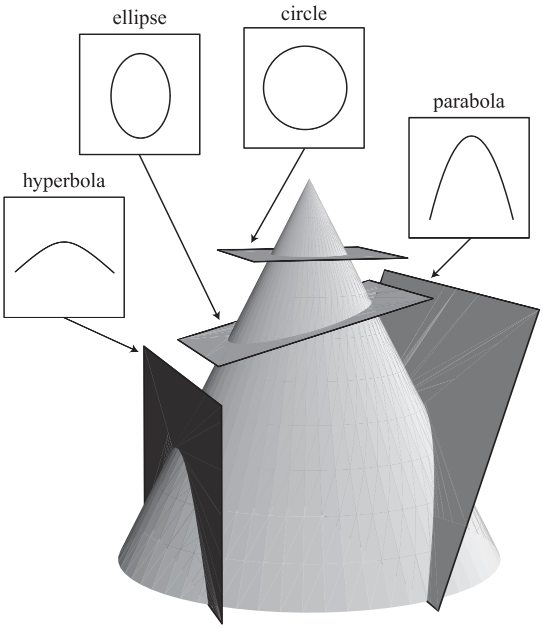


Figure 2.3 Conic sections. (Murray and Dermott 1999)

a straight line. For all orbits, the three orientation angles i, Ω and ω are defined as in the elliptical case.

Whereas the energy of an orbit is uniquely specified by its semimajor axis (eq. 2.22), the angular momentum also depends on the orbit's eccentricity:

$$|\mathbf{L}| = m\sqrt{GMa(1-e^2)}. \tag{2.26}$$

As with energy, the angular momentum of a circular orbit follows immediately from equation (2.21). For a given semimajor axis, a circular orbit contains the maximum possible amount of angular momentum (eq. 2.26). This occurs because when $r = a$ for an eccentric orbit, the magnitude of the velocity is the same as that for a circular orbit (by conservation of energy), but not all of this velocity is directed perpendicular to the line connecting the two bodies.

2.2 The Three-Body Problem

Gravity is not restricted to interactions between the Sun and the planets or individual planets and their satellites, but rather all bodies feel the gravitational force of one another. The motion of two mutually gravitating bodies is **completely integrable** (i.e., there exists one independent integral or constraint per degree of freedom), and the relative trajectories of the two bodies are given by simple conic sections, as discussed earlier. However, when more bodies are added to the system, additional constraints are needed to specify the motion; not enough integrals of motion are available, so the trajectories of even three gravitationally interacting bodies cannot be deduced analytically except in certain limiting cases. The general three-body problem is quite complex, and little progress can be made without resorting to numerical integrations. Fortunately, various approximations based on large differences between the masses of the bodies and nearly circular and co-planar orbits (which are quite accurate for most Solar System applications) simplify the problem sufficiently that some important analytic results may be obtained.

If one of the bodies is of negligible mass (e.g., a small asteroid, a ring particle or an artificial satellite), its effects on the other bodies may be ignored; the simpler system that results is called the **restricted three-body problem**, and the small body is referred to as a **test particle**. If the relative motion of the two massive particles is a circle, we refer to the situation as the **circular restricted three-body problem**. An alternative to the restricted three-body problem is **Hill's problem**, in which the mass of one of the bodies is much greater than the other two, but there is no restriction on the masses of the two small bodies relative to one another. An independent simplification is to assume that all three bodies travel within the same plane, the **planar three-body problem**. Various, but not all, combinations of these assumptions are possible.

Most of the results presented in this section are rigorously true only for the circular restricted three-body problem. However, they are valid to a good approximation for many configurations that exist in the Solar System.

2.2.1 Jacobi's Constant and Lagrangian Points

Our study of the three-body problem begins by considering an idealized system in which two massive bodies move on circular orbits about their common center of mass. A third body is introduced that is much less massive than the smaller of the first two, so that to good approximation, it has no effect on the orbits of the other bodies. Our analysis is performed in a noninertial frame that rotates about the z-axis at a rate equal to the orbital frequency of the two massive bodies. We choose units such that the distance between the two bodies, the sum of the masses and the gravitational constant are all equal to one; this implies that the angular frequency of the rotating frame also equals unity (Problem 2-6). The origin is given by the center of mass of the pair, and the two bodies remain fixed at points on the x-axis, $\mathbf{r}_1 = (-m_2/(m_1 + m_2), 0)$ and $\mathbf{r}_2 = (m_1/(m_1 + m_2), 0)$. By convention, $m_1 \geq m_2$; in most Solar System applications, $m_1 \gg m_2$. The (massless) test particle is located at $\mathbf{r}$, so $|\mathbf{r} - \mathbf{r}_i|$ is the distance from mass m_i to the test particle. The velocity of the test particle in the rotating frame is denoted by v.

By analyzing a modified energy integral in the rotating frame, Carl Jacobi deduced the following constant of motion for the circular restricted three-body problem:

$$C_J = x^2 + y^2 + \frac{2m_1}{|\mathbf{r} - \mathbf{r}_1|} + \frac{2m_2}{|\mathbf{r} - \mathbf{r}_2|} - v^2. \quad (2.27)$$

The first two terms on the right-hand side of equation (2.27) represent twice the centrifugal potential energy, the next two twice the gravitational potential energy and the final one twice the kinetic energy; C_J is known as **Jacobi's constant**. Note that a body located far from the two masses and moving slowly in the inertial frame has small C_J because the gravitational potential energy terms are small and the centrifugal potential almost exactly cancels the kinetic energy of the test particle's motion viewed in the rotating frame.

For a given value of Jacobi's constant, equation (2.27) specifies the magnitude of the test particle's velocity (in the rotating frame) as a function of position. Because v^2 cannot be negative, surfaces at which $v = 0$ bound the trajectory of a particle with fixed C_J (note that the allowed region need not be finite). Such **zero-velocity surfaces**, or in the case of the planar problem **zero-velocity curves**, are quite useful in discussing the topology of the circular restricted three-body problem.

Joseph Lagrange found that in the circular restricted three-body problem there are five points where test particles placed at rest would feel no net force in the rotating frame. The locations of three of these so-called **Lagrangian points** (L_1, L_2 and L_3) lie along a line joining the two masses m_1 and m_2. Zero-velocity curves intersect at each of the three collinear Lagrangian points, which are saddle points of the total (centrifugal + gravitational) potential in the rotating frame. The other two Lagrangian points (L_4 and L_5) form equilateral triangles with the two massive bodies. All five Lagrangian points are in the orbital plane of the two massive bodies. Figure 2.4 illustrates the positions of the Lagrangian points as well as trajectories and zero-velocity curves of various orbits that are close to these equilibrium positions.

Particles displaced slightly from the three collinear Lagrangian points will continue to move away; hence, these locations are unstable. The triangular Lagrangian points are potential energy maxima, but the Coriolis force stabilizes

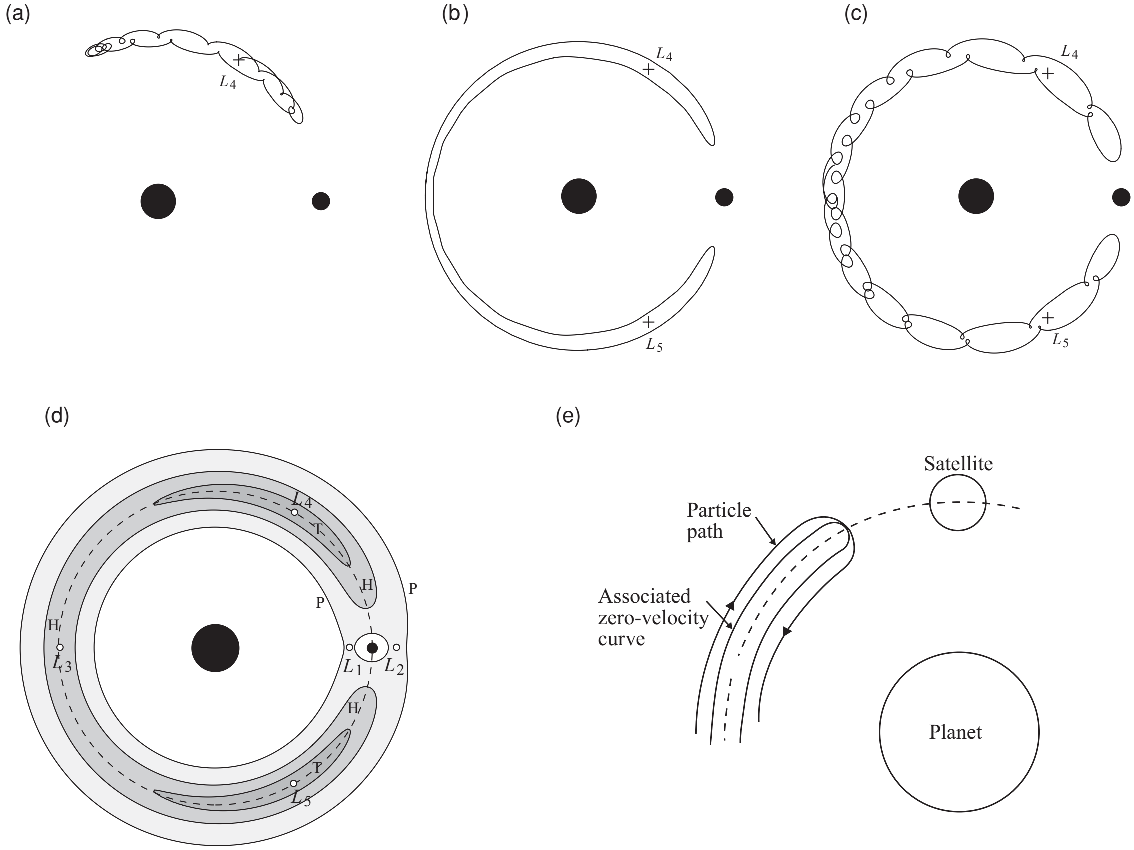


Figure 2.4 Schematic diagrams illustrating various properties of orbits in the circular restricted three-body problem. All cases are shown in the frame that is centered on the primary and rotating at the orbital frequency of the two massive bodies (corotating with the secondary). (a) Example of a tadpole orbit of a test particle viewed in the rotating frame. (b) Similar to (a) but for a horseshoe orbit with small eccentricity. (c) As in (b) but the particle has a larger eccentricity. (Panels a–c adapted from Murray and Dermott 1999) (d) The Lagrangian equilibrium points and various zero-velocity curves for three values of the Jacobi's constant, C_J. The mass ratio $m_1/m_2 = 100$. The locations of the Lagrangian equilibrium points L_1–L_5 are indicated by *small open circles*. The *white region* centered on the secondary is the secondary's Hill sphere. The *dashed line* denotes a circle of radius equal to the secondary's semimajor axis. The letters T (tadpole), H (horseshoe) and P (passing) denote the type of orbit associated with the curves. The regions enclosed by each curve (*shaded*) are excluded from the motion of a test particle that has the corresponding C_J. The largest horseshoe curve actually passes through L_2, and the largest tadpole curve passes through L_3. Horseshoe orbits can exist between these two extremes. (Courtesy Carl Murray) (e) Schematic diagram showing the relationship between a horseshoe orbit and its associated zero-velocity curve. The particle's velocity in the rotating frame drops as it approaches the zero-velocity curve, and it cannot cross the curve. (Adapted from Dermott and Murray 1981)

them for $m_1/m_2 \gtrsim 25$, which is the case for all known examples in the Solar System that are more massive than the Pluto–Charon system. If a particle at L_4 or L_5 is perturbed slightly, it will start to **librate** about these points (i.e., oscillate back and forth, without circulating past the secondary).

The L_4 and L_5 points are important in the Solar System. For example, the **Trojan asteroids** are located near Jupiter's triangular Lagrangian points, several asteroids are known to librate about Neptune's L_4 point, and several small asteroids, including 5261 Eureka, are martian Trojans. There are also small moons in the saturnian system near the triangular Lagrangian points of Tethys and Dione (Table E.4). The L_4 or L_5 points in the Earth–Moon system have been suggested as possible locations for a future space station.

2.2.2 Horseshoe and Tadpole Orbits

Consider a moon on a circular orbit about a planet. A particle just interior to the moon's orbit has a higher angular velocity and moves with respect to the moon in the direction of corotation. A particle just outside the moon's orbit has a smaller angular velocity and moves relative to the moon in the opposite direction. When the outer particle approaches the moon, the particle is pulled towards the moon and consequently loses angular momentum. Provided the initial difference in semimajor axis is not too large, the particle drops to an orbit lower than that of the moon. The particle then recedes in the forward direction. Similarly, the particle on the lower orbit is accelerated as it catches up with the moon, resulting in an outward motion towards a higher, and therefore slower, orbit. Orbits like these encircle the L_3, L_4 and L_5 points and appear shaped like horseshoes in the rotating frame (Fig. 2.4b); thus they are called **horseshoe orbits**. Saturn's small moons Janus and Epimetheus execute just such a dance, changing orbits every 4 years, as illustrated schematically in Figure 2.5. Because Janus and Epimetheus are comparable in mass, Hill's approximation is more accurate than is the restricted three-body formalism used earlier, but the dynamical interactions are essentially the same.

Because the Lagrangian points L_4 and L_5 are stable, material can librate about these points individually; such orbits are called **tadpole orbits** after their asymmetric elongated shape in the rotating frame (Fig. 2.4a). The Trojan asteroids librate about Jupiter's L_4 and L_5 points. The tadpole libration width at L_4 and L_5 is proportional to $(m_2/m_1)^{1/2}r$, and the horseshoe width varies as $(m_2/m_1)^{1/3}r$, where m_1 is the mass of the primary, m_2 the mass of the secondary and r the distance between the two objects. For a planet of Saturn's mass, $M_♄ = 5.7 \times 10^{26}$ kg, and a typical moon of mass $m_2 = 10^{17}$ kg (a 30-km-radius object with density of $\sim$1000 kg m^{-3}) at a distance of 2.5 R$_♄$, the tadpole libration half-width is $\sim$3 km and the horseshoe half-width $\sim$60 km.

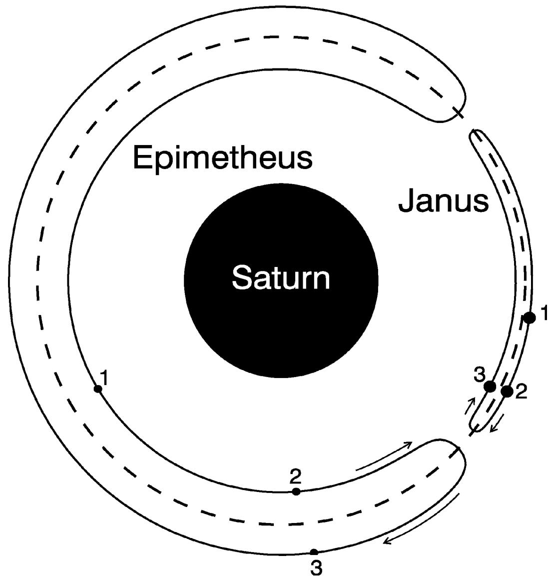


Figure 2.5 Diagram of the librational behavior of the Janus and Epimetheus co-orbital system in a frame rotating with the average mean motion of both satellites. The system is shown to scale, apart from the radial extent of the librational arcs being exaggerated by a factor of 500 and the radii of the moons inflated by a factor of 50. The ratio of the radial widths (as well as the azimuthal extents) of the arcs is equal to the Janus/Epimetheus mass ratio ($\sim$0.25). The numbered points represent a temporal sequence of positions of the two moons over approximately one-fourth of a libration cycle. (Tiscareno et al. 2009)

2.2.3 Hill Sphere

The approximate limit to a secondary's (e.g., planet's or moon's) gravitational dominance is

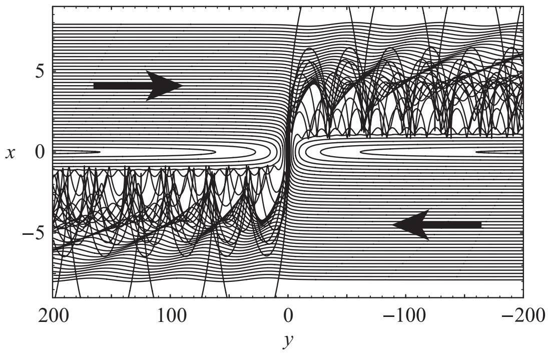


Figure 2.6 The trajectories of 80 test particles in the vicinity of a secondary of mass $m_2 \ll m_1$ are shown in the frame rotating with the secondary's (circular) orbit about the primary. The scale of the plot is expanded in the radial (x) direction relative to that in the azimuthal (y) direction, with numerical values in both directions given in units of the radius of the secondary's Hill sphere. The secondary mass is located at the origin and the L_1 and L_2 points are at $y = 0$, $x = \pm 1$. The particles were all started with $dx/dt = 0$ (i.e., circular orbits) at $y = \pm 200$. The *arrows* indicate their direction of motion before encountering the secondary. The primary is located at $y = 0, x = -\infty$. In an inertial frame, the secondary and the test particles all move from right to left. (Adapted from Murray and Dermott 1999)

given by the extent of its **Hill sphere**,

$$R_H = \left(\frac{m_2}{3(m_1 + m_2)} \right)^{1/3} a, \tag{2.28}$$

where m_2 is the mass of the secondary and m_1 the primary's (e.g., Sun's or planet's) mass. The Hill sphere stretches out to the L_1 point and essentially circumscribes the Roche lobe (§13.1) in the limit $m_2 \ll m_1$. Planetocentric orbits that are stable over long periods of time are those well within the boundary of a planet's Hill sphere; all known natural satellites lie in this region. As illustrated in Figure 2.6, stable heliocentric orbits are always well outside the Hill sphere of any planet. Comets and other bodies that enter the Hill sphere of a planet at very low velocity can remain gravitationally bound to the planet for some time as **temporary satellites**, an example of which is shown in Figure 2.7.

The orbits of moons that lie in the inner part of a planet's Hill sphere are classified as prograde if the moons move in the sense that the planet rotates and retrograde if they travel in the opposite sense.

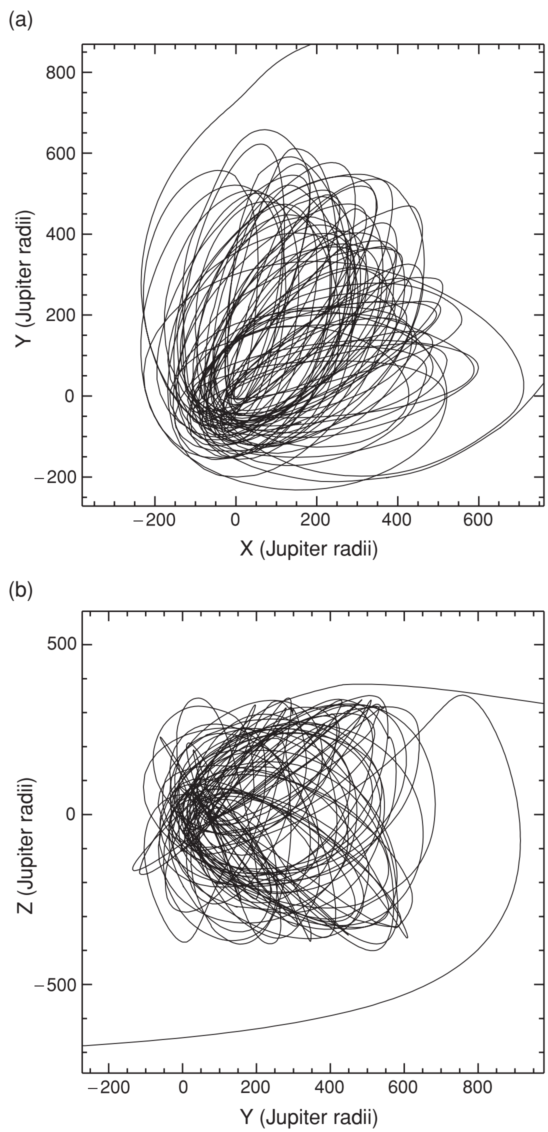


Figure 2.7 Trajectory relative to Jupiter of a test particle initially orbiting the Sun that was temporarily captured into an unusually long duration (140 years) unstable orbit about Jupiter. (a) Projected into the plane of Jupiter's orbit about the Sun. (b) Projected into a plane perpendicular to Jupiter's orbit. (Kary and Dones 1996)

However, for very distant satellites, the more important dynamical criterion is whether they travel in the same direction as the planet orbits the Sun (prograde) or in the opposite sense (retrograde). Retrograde orbits are stable to larger distances from a planet than are prograde ones, and moons on retrograde orbits are found at greater distances (Table E.4).

2.3 Perturbations and Resonances

Within the Solar System, one body typically produces the dominant gravitational force on any given object, and the resultant motion can be thought of as a Keplerian orbit about a primary, subject to small perturbations by other bodies. Although perturbations on a body's orbit are often small, they cannot always be ignored. They must be included in short-term calculations if high accuracy is required, e.g., for predicting stellar occultations or targeting spacecraft. Most long-term perturbations are periodic in nature, their directions oscillating with the relative longitudes of the bodies or with some more complicated function of the bodies' orbital elements. Small perturbations can produce large effects if the forcing frequency is commensurate or nearly commensurate with the natural frequency of oscillation of the responding elements. Under such circumstances, perturbations add coherently, and the effects of many small tugs can build up over time to create a large-amplitude, long-period response. This is an example of **resonance forcing**, which occurs in a wide range of physical systems. In this section, we consider some important examples of the effects of these perturbations on the orbital motion.

2.3.1 Resonant Forcing

An elementary example of resonance forcing is given by the one-dimensional forced harmonic oscillator, for which the equation of motion is

$$m\frac{d^2x}{dt^2} + m\omega_o^2 x = F_f \cos \omega_f t, \tag{2.29}$$

where m is the mass of the oscillating particle, F_f is the amplitude of the driving force, ω_o is the natural frequency of the oscillator and ω_f is the forcing frequency. The solution to equation (2.29) is

$$x = \frac{F_f}{m\left(\omega_o^2 - \omega_f^2\right)} \cos \omega_f t + C_1 \cos \omega_o t + C_2 \sin \omega_o t, \tag{2.30}$$

where C_1 and C_2 are constants determined by the initial conditions. Note that if $\omega_f \approx \omega_o$, a large-amplitude, long-period response can occur even if F_f is small. Moreover, if $\omega_o = \omega_f$, equation (2.30) is invalid. In this (resonant) case, the solution is given by

$$x = \frac{F_f}{2m\omega_o} t \sin \omega_o t + C_1 \cos \omega_o t + C_2 \sin \omega_o t. \tag{2.31}$$

The t in the middle of the first term at the right-hand side of equation (2.31) leads to secular (i.e., steady rather than periodic) growth. Often this linear growth is moderated by the effects of nonlinear terms that are not included in the simple example provided above. However, some perturbations have a secular component.

2.3.2 Mean Motion Resonances

The simplest celestial resonances to visualize are so-called **mean motion resonances**, in which the orbital periods of two bodies are commensurate (e.g., have a ratio of the form $N/(N+1)$ or $N/(N+2)$, where N is an integer). Some examples of the consequences of mean motion resonance are given below. Almost exact orbital commensurabilities exist at many places in the Solar System. As illustrated in Figure 2.8, Io orbits Jupiter twice as frequently as Europa does, and Europa in turn orbits Jupiter in half of the time that Ganymede

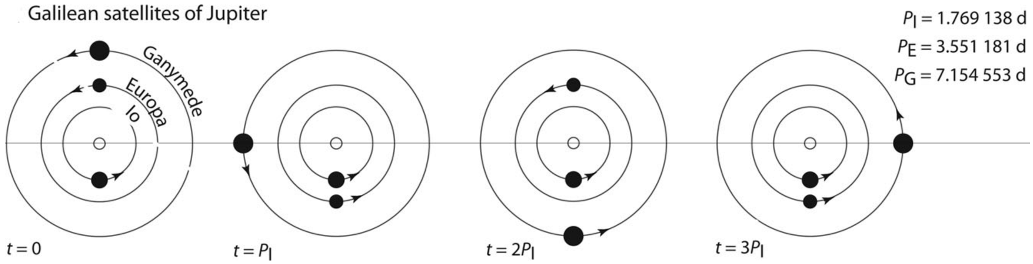


Figure 2.8 Schematic illustration of the orbital resonances between the three inner Galilean satellites of Jupiter. Successive views represent the system at times separated by one orbital period of the moon Io, P_I. The configuration at 4 P_I is identical to that at $t = 0$. (From Perryman 2011)

takes. **Conjunction** (the moons being at the same longitude in their orbits about the planet) between Io and Europa always occurs when Io is at its perijove (the point in its orbit that is closest to Jupiter). How can such commensurabilities exist? After all, the rational numbers form a set of measure zero on the real line, which means that the probability of randomly picking a rational from the real number line is nil! The answer lies in the fact that **orbital resonances** may be held in place by stable 'locks' that result from nonlinear effects not represented in the simple mathematical example of the harmonic oscillator. Differential tidal recession (§2.7.2) brings moons into resonance, and nonlinear interactions between the moons can keep them there.

Other examples of resonance locks include the Hilda and Trojan asteroids with Jupiter, Neptune–Pluto, and several pairs of moons orbiting Saturn, such as Janus–Epimetheus, Mimas–Tethys and Enceladus–Dione. Resonant perturbations can force bodies into eccentric and/or inclined orbits, which may lead to collisions with other bodies; this is thought to be the dominant mechanism for clearing the Kirkwood gaps in the asteroid belt (see below). Several moons of Jupiter and Saturn have significant resonantly produced **forced eccentricities**, which are denoted by the symbol f in Table E.4.

Spiral waves can be produced in a self-gravitating disk of particles by resonant perturbations of a satellite. Spiral density waves resonantly excited by moons are observed in Saturn's rings (§13.4.2). Analogous waves in protoplanetary disks can alter the orbits of young planets (§15.7.1).

2.3.3 Secular Resonances

Although many interactions among planets depend on their relative azimuthal positions, some important long-term effects are produced by the shapes and orientations of their orbits (eccentricities, apse locations, inclinations and nodes). **Secular perturbation theory** averages over orbital timescales and treats planets as elliptical wires of nonuniform density, with density corresponding to the time spent at a given phase of the orbit according to Kepler's second law (§2.1.1). Secular perturbations can alter the eccentricity and inclination of an orbit but not its semimajor axis. A **secular resonance** occurs when the apses or nodes of two orbits precess at the same rate.

In the restricted circular three-body problem, secular perturbations can change the small body's eccentricity and its inclination relative to the orbit of the two massive bodies, but the quantity $\sqrt{1 - e^2} \cos i$ remains constant. Orbital inclination can thus be traded for eccentricity. For high values of inclination, $\cos^2 i < 3/5$, the **Kozai mechanism** forces the argument of periapse to remain fixed, and large periodic variations in eccentricity and inclination are produced. The Kozai mechanism causes some asteroids and comets to approach closely to and even collide with the Sun (§12.3.5) and highly inclined irregular satellites to collide with their planets. It is also likely to be one of the mechanisms responsible for the high observed eccentricities of some extrasolar planets (Fig. 14.22) and the high inclinations some of these planets' orbits have relative to the planes of their star's equator (§14.2.3).

2.3.4 Resonances in the Asteroid Belt

There are obvious patterns in the distribution of asteroidal semimajor axes that appear to be associated with mean motion resonances with Jupiter (Fig. 12.2a). At these resonances, a particle's period of revolution about the Sun is a small integer ratio multiplied by Jupiter's orbital period. The Trojan asteroids travel in a 1:1 mean motion resonance with Jupiter. Trojan asteroids execute small-amplitude (tadpole) librations about the L_4 and L_5 points 60° behind or ahead of Jupiter and therefore never have a close approach to Jupiter. Another example of a protection mechanism provided by a resonance is the Hilda group of asteroids at Jupiter's 3:2 mean motion resonance and the asteroid 279 Thule at the 4:3 resonance. The Hilda asteroids have a libration about 0° of their **critical argument** (the combination of orbital elements that signifies the resonant configuration), $3\lambda' - 2\lambda - \varpi$, where λ' is Jupiter's longitude, λ is the asteroid's longitude and ϖ is the asteroid's longitude of perihelion. In this way, whenever the asteroid is in conjunction with Jupiter ($\lambda = \lambda'$), the asteroid is close to perihelion ($\lambda' \approx \varpi$) and well away from Jupiter.

Most orbits starting with small eccentricity in the general vicinity of the 3:1 mean motion resonance with Jupiter appear regular and show very little variation in eccentricity or semimajor axis over timescales of 5×10^4 yrs. However, orbits near the resonance can maintain a low eccentricity ($e < 0.1$) for nearly a million years and then have a 'sudden' increase in eccentricity to $e > 0.3$. Asteroids that begin on near-circular orbits in the gap acquire sufficient eccentricities to cross the orbits of Mars and the Earth and in some cases become so eccentric that they hit the Sun, so the perturbative effects of the terrestrial planets are probably capable of clearing out the 3:1 gap in a time equivalent to the age of the Solar System.

The ν_6 secular resonance occurs where the periapse angle of an asteroid precesses at the rate of the sixth secular frequency of our Solar System, which is essentially the same as the precession rate of Saturn's periapse. Perturbations resulting from the ν_6 resonance can excite asteroidal eccentricities to such high values that the ν_6 resonance is largely responsible for the inner edge of the asteroid belt near 2.1 AU.

2.3.5 Regular and Chaotic Motion

Direct integrations of multi-body systems on computers demonstrate that for some initial conditions, the trajectories are **regular** with variations in their orbital elements that seem to be well described by the perturbation series, but for other initial conditions, the trajectories are found to be **chaotic** and are not as confined in their motions. The evolution of a system that is chaotic depends so sensitively on the system's precise initial state that the behavior is in effect unpredictable even though it is strictly determinate in a mathematical sense.

Figure 2.9 shows a key feature of chaotic orbits that we use here as a definition of **chaos**: Two trajectories that begin arbitrarily close in phase space (which can be defined using coordinates such as positions and velocities, or a more complicated set of orbital elements) within a chaotic region typically diverge exponentially in time. Within a given chaotic region, the timescale for this divergence does not typically depend on the precise values of the initial conditions! The distance, $d(t)$, between two particles having an initially small separation, $d(0)$, increases slowly for regular orbits,

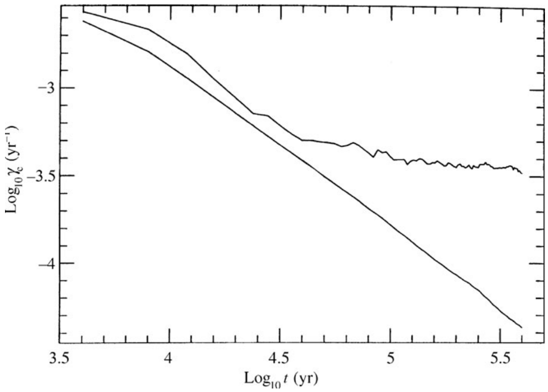


Figure 2.9 Distinction between regular (*lower curve*, nearly straight) and chaotic trajectories (*upper curve*) as characterized by the Lyapunov characteristic exponent, γ_c. Both trajectories are near the 3:1 resonance with Jupiter, and they have been integrated using the elliptic restricted three-body problem. For chaotic trajectories, a plot of $\log \gamma_c$ versus $\log t$ eventually levels off at a value of γ_c that is the inverse of the Lyapunov timescale for the divergence of initially adjacent trajectories. For regular trajectories, $\gamma_c \to 0$ as $t \to \infty$. (Adapted from Duncan and Quinn 1993)

with $d(t) - d(0)$ growing as a power of time t (typically linearly). In contrast, for chaotic orbits,

$$d(t) \sim d(0)e^{\gamma_c t}, \tag{2.32}$$

where γ_c is the **Lyapunov characteristic exponent** and γ_c^{-1} is the **Lyapunov timescale**. From this definition of chaos, we see that chaotic orbits show such a sensitive dependence on initial conditions that the detailed long-term behavior of the orbits is lost within several Lyapunov timescales. Even a fractional perturbation as small as 10^{-10} in the initial conditions will result in a 100% discrepancy in about 20 Lyapunov times. However, one of the interesting features of much of the chaotic behavior seen in simulations of the orbital evolution of bodies in the Solar System is that the timescale for large changes in the principal orbital elements is often many orders of magnitude longer than the Lyapunov timescale.

In dynamical systems such as the Solar System, chaotic regions do not appear randomly; rather, many of them are associated with trajectories in which the ratios of characteristic frequencies of the original problem are sufficiently well approximated by rational numbers, i.e., near resonances. Figure 2.10 shows that the outer boundaries of the chaotic zone coincide well with the boundaries of the 3:1 Kirkwood gap.

The above discussion applies to orbits that do not closely approach any massive secondaries. Close approaches can lead to highly chaotic and unpredictable orbits, such as the possible future behaviors of the giant, distant cometary centaur Chiron shown in Figure 2.11 (see §12.2.2.). These planet-crossing trajectories do not require resonances to be unstable and generally are not well characterized by a constant Lyapunov exponent.

For nearly circular and coplanar orbits, the strongest mean motion resonances occur at locations where the ratio of test particle orbital periods to the massive body's period is of the form

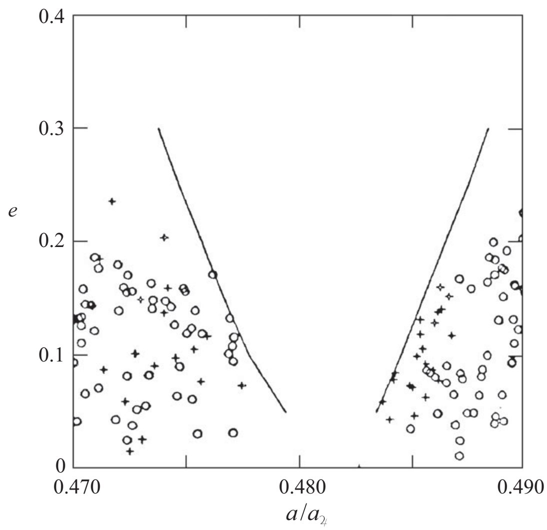


Figure 2.10 The outer boundaries of the chaotic zone surrounding Jupiter's 3:1 mean motion resonance in the a–e plane are shown as *lines*. Locations of numbered asteroids are shown as *circles* and Palomar–Leiden survey asteroids (whose orbits are less well determined) are represented as *plus signs*. Note the excellent correspondence of the observed 3:1 Kirkwood gap with theoretical predictions. (Adapted from Wisdom 1983)

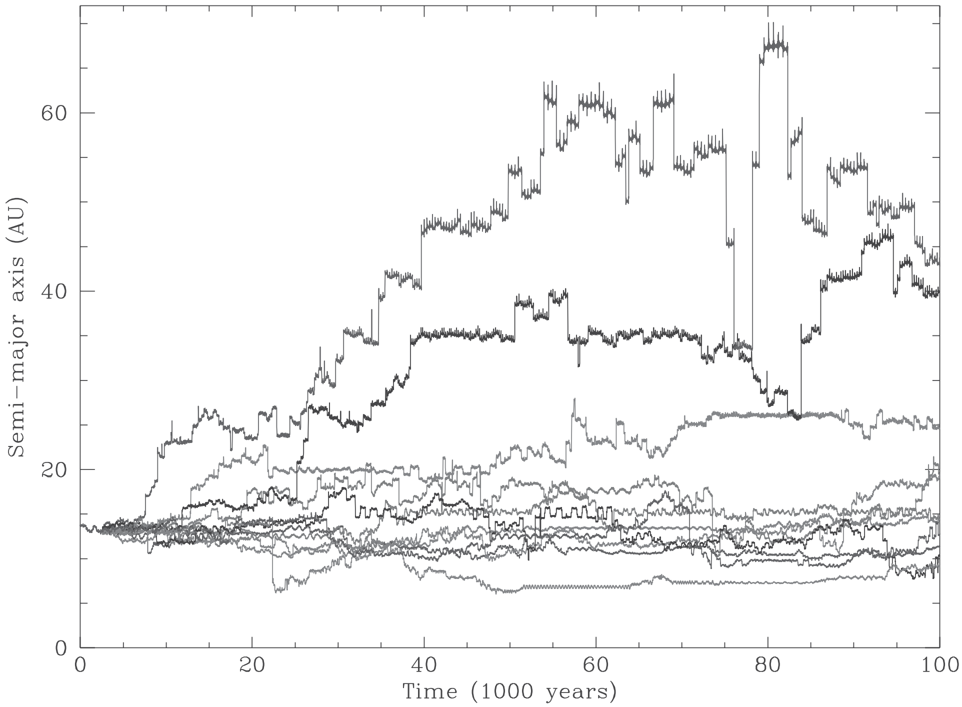


Figure 2.11 COLOR PLATE The future evolution of the semimajor axis of P/Chiron's orbit according to 11 numerical integrations. The initial orbital elements of the simulated bodies differed by about 1 part in 10^6. The orbit of Chiron currently crosses the orbits of both Saturn and Uranus, and is not protected from close approaches with either planet by any resonance. Chiron's orbit is highly chaotic, with gross divergence of trajectories in $<10^4$ years. (Courtesy L. Dones)

N:($N \pm 1$), where N is an integer. At these locations, conjunctions (closest approaches) always occur at the same phase in the orbit, and tugs add coherently. (The locations of these strong resonances are shifted slightly when the primary is oblate; see §§2.6 and 13.4 for details.) The strength of these first-order resonances increases as N grows because the magnitudes of the perturbations are larger closer to the secondary. First-order resonances also become closer to one another near the orbit of the secondary (Problem 2-7). Sufficiently close to the secondary, the combined effects of greater strength and smaller spacing cause resonance regions to overlap; this overlapping can lead to the onset of chaos as particles shift between the nonlinear perturbations of various resonances. The region of overlapping resonances is approximately symmetric about the planet's orbit and has a half-width, Δa_{ro}, given by

$$\Delta a_{ro} \approx 1.5 \left(\frac{m_2}{m_1}\right)^{2/7} a, \qquad (2.33)$$

where a is the semimajor axis of the planet's orbit. Whereas the functional form of equation (2.33) has been derived analytically, the coefficient 1.5 is a numerical result.

2.4 Stability of the Solar System

We turn now to one of the oldest problems in dynamical astronomy: whether or not the planets will continue indefinitely in almost circular, almost

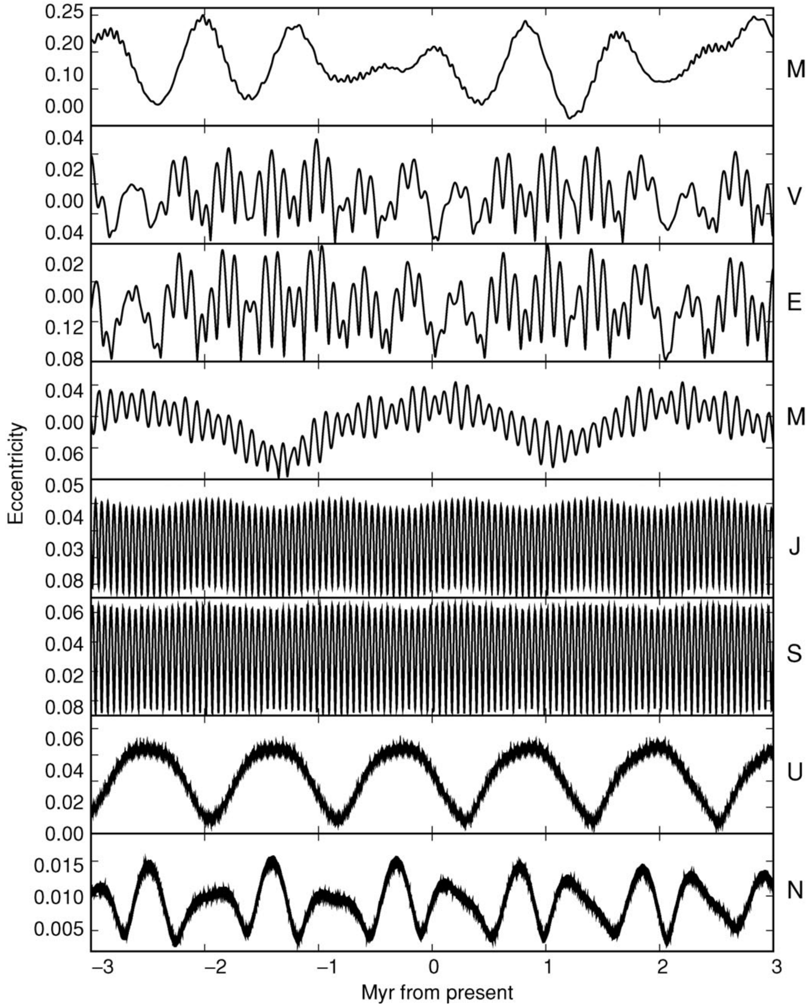


Figure 2.12 The eccentricities of the eight major planets are shown for 6×10^6 years centered on the present epoch. Mercury's eccentricity is displayed in the *top panel* followed by that of each of the other planets in order of their heliocentric distance. Note the relatively large amplitudes of the variations of the two smallest planets, Mercury and Mars, and the correlated oscillations of $e_{\oplus}$ with those of $e_{♀}$ and of $e_{♃}$ with those of $e_{♄}$. (Courtesy Tom Quinn; see Laskar et al. 1992 for an explanation of the integration used to compute these values)

coplanar orbits. From an astronomical viewpoint, stability implies that the system will remain bound (no ejections), that no mergers of planets will occur for the possibly long but finite period of interest and that this result is robust against (most if not all) sufficiently small perturbations.

2.4.1 Orbits of the Eight Planets

Figure 2.12 shows the behavior of the eccentricities of all eight planets for 3 million years into the past as well as into the future. Mercury's eccentricity reaches higher values on 10^8-year timescales, as can be seen in Figure 2.13, but the eccentricities of the other planets do not extend much beyond their range shown in Figure 2.12 over this time interval. Variations in the semimajor axis of Earth's orbit over ± 3 Myr are shown in Figure 2.14. The small fractional changes in semimajor axis relative to the variations in eccentricity evident for Earth are characteristic of all eight planets.

Long-duration numerical integrations show a surprisingly high Lyapunov exponent, $\sim(5\ \mathrm{Myr})^{-1}$. Such large Lyapunov exponents certainly suggest chaotic behavior. However, the

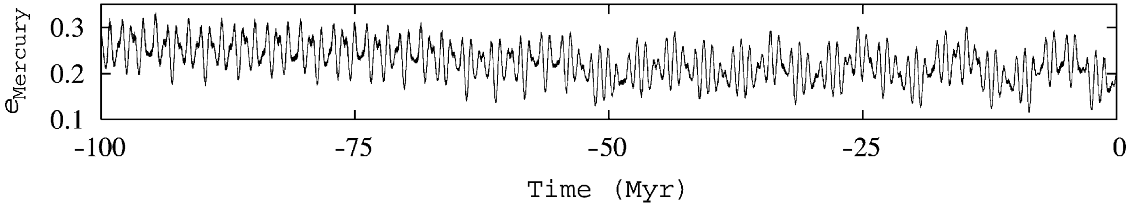


Figure 2.13 Variations in the eccentricity of Mercury's orbit over the past 100 million years. Integrations included the Sun, all eight planets and first-order post-Newtonian effects of general relativity; the eccentricities of all of the planets over the past 3 million years look the same in the integration used to produce this figure as those shown in Figure 2.12. (Courtesy Julie Gayon)

apparent regularity of the motion of the Earth, and indeed the fact that the Solar System has survived for 4.5 billion years, implies that any pathways through phase space that might lead to (highly chaotic) close approaches must be narrow. Nonetheless, the exponential divergence seen in all long-term integrations implies that the accuracy of the deterministic equations of celestial mechanics to predict the future positions of the planets will always be limited by the accuracy with which their orbits can be measured. For example, even if the position of Earth along its orbit were to be known to within 1 cm today and all other planetary masses, positions and velocities were known exactly, the exponential propagation of errors that is characteristic of chaotic motion implies that we would still have no knowledge of Earth's orbital longitude 200 million years in the future.

The situation is even less predictable when the gravitational influence of smaller bodies is accounted for. Asteroids exert small perturbations on the orbits of the major planets. These perturbations can be accounted for and do not adversely affect the precision to which planetary orbits can be simulated on timescales of tens of millions of

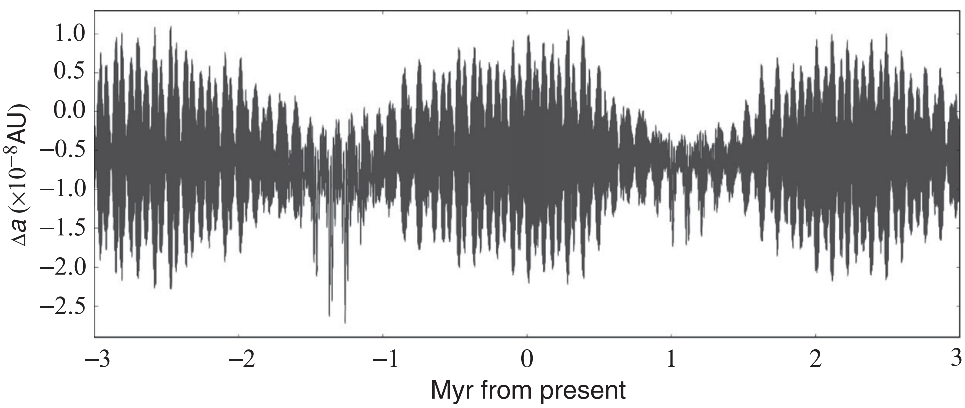


Figure 2.14 Variations in the semimajor axis of Earth's orbit (more precisely, the semimajor axis of the center of mass of the Earth–Moon system) over a time interval of 6 million years centered on the present epoch. Note the scale of the vertical axis, which indicates that the Earth's semimajor axis varies by only a few kilometers over timescales of millions of years. These data were taken from the integrations used to produce the plot of eccentricities of the planets shown in Figure 2.12. (Courtesy Tom Quinn)

years. However, unlike the major planets, asteroids suffer close approaches to one another and thus are subject to the types of chaos depicted in Figure 2.11. Close approaches between the two largest and most massive asteroids, 1 Ceres and 4 Vesta (Tables E.6 and E.8), lead to exponential growth in uncertainty for backwards integrations of planetary orbits with doubling times of $<10^6$ years prior to 50–60 million years ago.

It is also worth bearing in mind the lessons learned from integration of test particle trajectories, namely that the timescale for macroscopic changes in the system can be many orders of magnitude longer than the Lyapunov timescales. Thus, the apparent stability of the current planetary system on billion-year timescales may simply be a manifestation of the fact that the Solar System is in the chaotic sense a dynamically young system. Indeed, there is a small but nontrivial chance that Mercury's orbit will become so eccentric that it will cross the orbit of Venus before the Sun evolves off the main sequence 6 billion years from now.

Because planetary perturbations appear to be capable of bringing the Solar System to the verge of instability on geological timescales, the planets within our Solar System may be about as closely spaced as can be expected for a mature planetary system containing planets as massive as those orbiting the Sun. Although somewhat more crowded configurations can be long lived, it may well be that the planet formation process (see Chapter 15) is unlikely to produce more densely packed systems of similar planets that survive on gigayear timescales.

2.4.2 Survival Lifetimes of Small Bodies

Interplanetary space is vast, yet few bodies orbit within this great expanse. And those few bodies are far from randomly distributed. Rather, minor planets are concentrated within a few regions (§12.2): the Kuiper belt beyond Neptune's orbit, the main asteroid belt between the orbits of Mars and Jupiter, the regions surrounding the triangular Lagrangian points of the Sun–Jupiter system (§2.2.1), and probably around the regions surrounding the triangular Lagrangian points of the Sun–Neptune system. Dynamical analyses show that orbits within these regions remain stable for far longer than trajectories passing through most other locations in the Solar System. What causes the removal of bodies from other regions of the Solar System? How rapidly are they removed?

Trajectories crossing the paths of one or more of the major planets are rapidly destabilized by scatterings resulting from close planetary approaches unless they are protected by some type of resonance (as is Pluto). Small bodies can remain on orbits between a pair of terrestrial planets or a pair of giant planets for much longer, but most are perturbed into planet-crossing paths in less than the age of the Solar System by the same resonance overlap-induced chaos that makes planetary orbits unpredictable on long timescales. Lifetimes of orbits vary greatly, and collections of test particles spread randomly over even fairly small regions of phase space last for quite diverse amounts of time. Figure 2.15 illustrates stability times for test particles located exterior to 5 AU; note the logarithmic scale for the time axis. Loss rates are rapid early on, but as particles near the stronger resonances are removed, it takes longer and longer for a given fraction of the remaining bodies to be destabilized. This decay rate is more gradual than that of other natural processes, such as radioactivity (§3.4), in which the population drops exponentially with time (§11.6.1).

2.5* Dynamics of Spherical Bodies

Thus far we have approximated Solar System bodies as point masses for the purpose of calculating their mutual gravitational interactions. Self-gravity

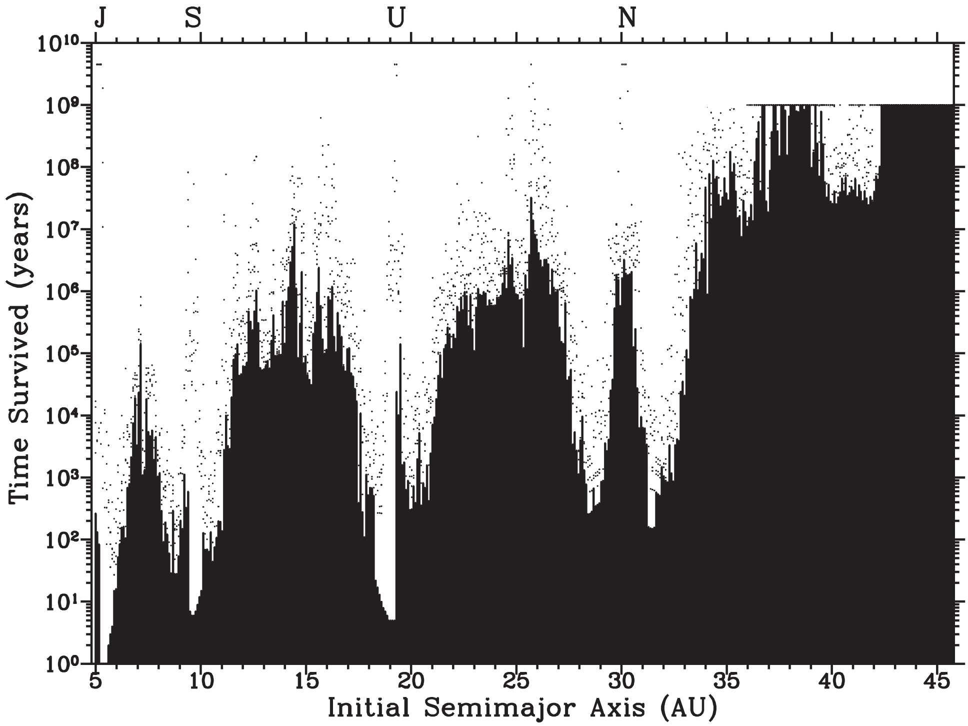


Figure 2.15 Stability map for test particles in the outer Solar System based on numerical integrations that include the Sun and the four giant planets. The time that each particle survived is plotted as a function of particle initial semimajor axis. For each semimajor axis bin, six particles were started at differing longitudes. The *vertical bars* mark the minimum of the six termination times. The *points* mark the termination times of the other five particles. The *scatter of points* gives an idea of the spread in particle lifetimes at each semimajor axis. The locations of the planets are denoted on the top of the figure; the spikes in particle lifetimes near these semimajor axes represent particles initially in tadpole or horseshoe orbits. The integrations extend to 4.5×10^9 years for particles initially interior to Neptune and to 10^9 years for those farther out. Only a few particles initially interior to Neptune survived the entire integrations, but many particles exterior to 33 AU and all particles beyond about 43 AU remained on non–planet-crossing orbits for the entire time interval simulated. (Courtesy Matt Holman; see Holman 1997 for details on the calculations)

causes most large celestial bodies to be approximately spherically symmetric.

2.5.1 Moment of Inertia

The moment of inertia, I, of a body about a particular axis is defined as

$$I \equiv \iiint \rho(\mathbf{r}) r_c^2 d\mathbf{r}, \tag{2.34}$$

where r_c is the distance from the axis and the integral is taken over the entire body. The rotational angular momentum of a simply rotating rigid body is given by

$$\begin{aligned} \mathbf{L} &= \iiint \rho(\mathbf{r})(\mathbf{r} \times \mathbf{v})\, d\mathbf{r} \\ &= \iiint \rho(\mathbf{r}) r_c^2 \omega_{rot} d\mathbf{r} = I\boldsymbol{\omega}_{rot}, \end{aligned} \tag{2.35}$$

where $\boldsymbol{\omega}_{rot}$ is its spin angular velocity. Orbital angular momentum is given by equations (2.12) and (2.26). Analogously, the kinetic energy of rotation is given by

$$E_{rot} = \frac{1}{2} I \omega_{rot}^2 = \frac{1}{2} \omega_{rot} L. \tag{2.36}$$

The moment of inertia of a uniform density sphere of radius R and mass m about its center of mass can be computed directly by performing the integration specified in equation (2.34) over the sphere. However, a more elegant and less tedious method exploits the symmetry of the sphere as follows. Note that to compute I about the z-axis, we have $r_c^2 = x^2 + y^2$; about the x-axis, $r_c^2 = y^2 + z^2$; and about the y-axis, $r_c^2 = x^2 + z^2$. By symmetry, these three integrals are equal, so adding them gives:

$$3I = \iiint \rho(2x^2 + 2y^2 + 2z^2)d\mathbf{r}. \tag{2.37}$$

The quantity in parentheses in equation (2.37) is equal to $2r^2$, so the integral can most easily be performed using spherical coordinates. Dividing both sides by 3 and noting that the integrations in the angular coordinates give the surface area of a spherical shell, $4\pi r^2$, yields:

$$I = \frac{8\pi\rho}{3}\int_0^R r^4 dr = \frac{2}{5}mR^2. \tag{2.38}$$

Centrally condensed bodies have moment of inertia ratios $I/(mR^2) < 2/5$ (Problem 2-10). The moment of inertia ratios for the planets are listed in Table E.15.

2.5.2 Gravitational Interactions

Newton showed that the gravitational force exerted by a spherically symmetric body exterior to its surface is identical to the gravitational force of a pointlike particle of the same mass located at the body's center. We derive this result below using multiple integration of the gravitational potential, which is defined in the next paragraph.

For many applications, it is convenient to express the gravitational field in terms of a potential, $\Phi_g(\mathbf{r})$, defined as:

$$\Phi_g(\mathbf{r}) \equiv -\int_\infty^{\mathbf{r}} \frac{\mathbf{F}_g(\mathbf{r}')}{m} \cdot d\mathbf{r}'. \tag{2.39}$$

By inverting equation (2.39), one can see that the gravitational force is the gradient of the potential and

$$\frac{d^2\mathbf{r}}{dt^2} = -\nabla\Phi_g. \tag{2.40}$$

In general, $\Phi_g(\mathbf{r})$ satisfies Poisson's equation:

$$\nabla^2\Phi_g = 4\pi\rho G. \tag{2.41a}$$

In empty space, $\rho = 0$, so $\Phi_g(\mathbf{r})$ satisfies Laplace's equation:

$$\nabla^2\Phi_g = 0. \tag{2.41b}$$

A spherically symmetric body can be viewed as the sum of thin concentric shells, each of uniform surface density. Without loss of generality, consider a shell of radius R and surface density unity centered at the origin and evaluate the potential at a location $\mathbf{r}$, where $r > R$. Subdivide the shell into rings that are oriented perpendicular to the direction from the center of the sphere to the point at which the potential is being evaluated. As illustrated in Figure 2.16, let θ denote the angle between lines from the origin to a point on the ring and to $\mathbf{r}$. The mass of the ring is given by $2\pi R \sin\theta$. The potential of the shell at the point under consideration is given by:

$$\Phi_g = \int_0^\pi 2\pi R \sin\theta \frac{1}{r'} R d\theta, \tag{2.42}$$

where r' denotes the distance from the point at which the potential is being evaluated. The law of cosines gives the square of r' as:

$$r'^2 = R^2 + r^2 - 2rR\cos\theta. \tag{2.43}$$

Because r and R are constant, we may differentiate equation (2.43) and rearrange the terms to obtain:

$$\frac{\sin\theta}{r'}d\theta = \frac{dr'}{Rr}. \tag{2.44}$$

Substituting equation (2.44) into equation (2.42) yields:

$$\Phi_g = 2\pi R^2 \int_{\mathrm{r-R}}^{r+R} \frac{dr'}{Rr} = \frac{4\pi R^2}{r}. \tag{2.45}$$

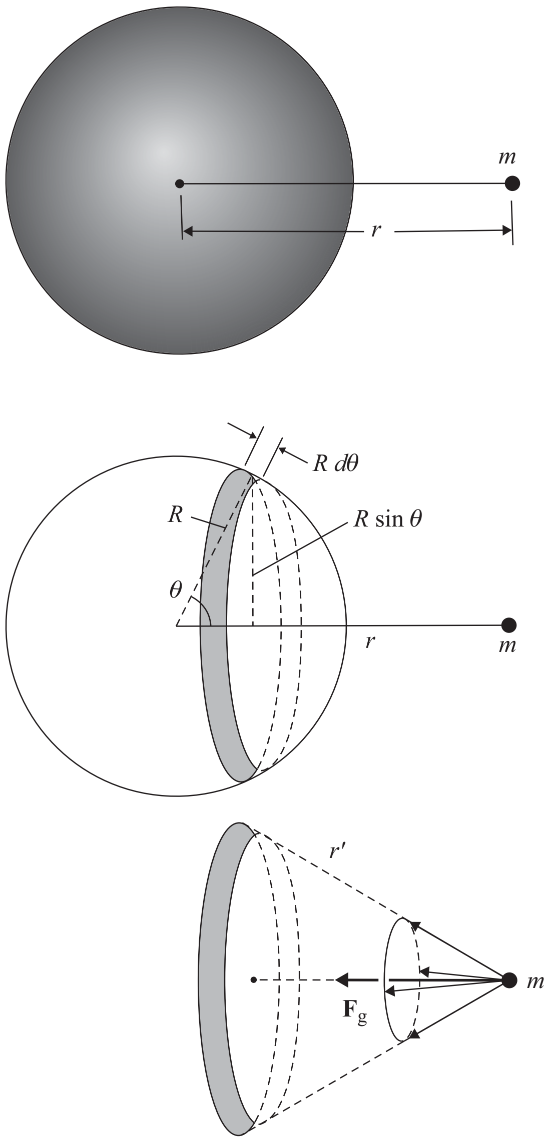


Figure 2.16 Diagram showing the notation used in the calculation of the gravitational potential exterior to a uniform sphere.

The value of Φ_g given in equation (2.45) is equal to the area of the sphere (and thus to its mass if surface density of unity is assumed) divided by the distance from its center to the point at which the potential is being evaluated. This value is identical to the potential of a point particle of the same mass located at the center of the sphere, completing the proof.

2.6 Orbits about an Oblate Planet

Several forces act to produce distributions of mass that deviate from spherical symmetry. In the Solar System, rotation, physical strength and tides produce important departures from spherical symmetry in some bodies. The gravitational field near an aspherical body differs from that near a point-mass. The gravity field can be determined to quite high accuracy by tracking the orbits of spacecraft close to the body or from the rate of precession of the periapses of moons and rings orbiting the planet. The gravity field of a planet or moon contains information on the body's internal density structure.

Most planets are very nearly axisymmetric, with the major departure from sphericity being due to a rotationally induced equatorial bulge. Thus, in this section, we analyze the effects of an axisymmetric body's deviation from spherical symmetry on the gravitational force that it exerts and on its response to external torques.

2.6.1* Gravity Field

The analysis of the gravitational field of an axisymmetric planet is most conveniently done by using the Newtonian gravitational potential, $\Phi_g(\mathbf{r})$, which is defined in equation (2.39). Because $\Phi_g(\mathbf{r})$ in free space satisfies Laplace's equation (2.41*b*), the gravitational potential exterior to a planet can be expanded in terms of **Legendre polynomials**:

$$\Phi_g(r,\phi,\theta) = -\frac{Gm}{r}\left[1 - \sum_{n=2}^{\infty} J_n P_n(\cos\theta)\left(\frac{R}{r}\right)^n\right]. \tag{2.46}$$

Equation (2.46) is written in standard spherical coordinates, with ϕ the longitude and θ representing the angle between the planet's symmetry axis