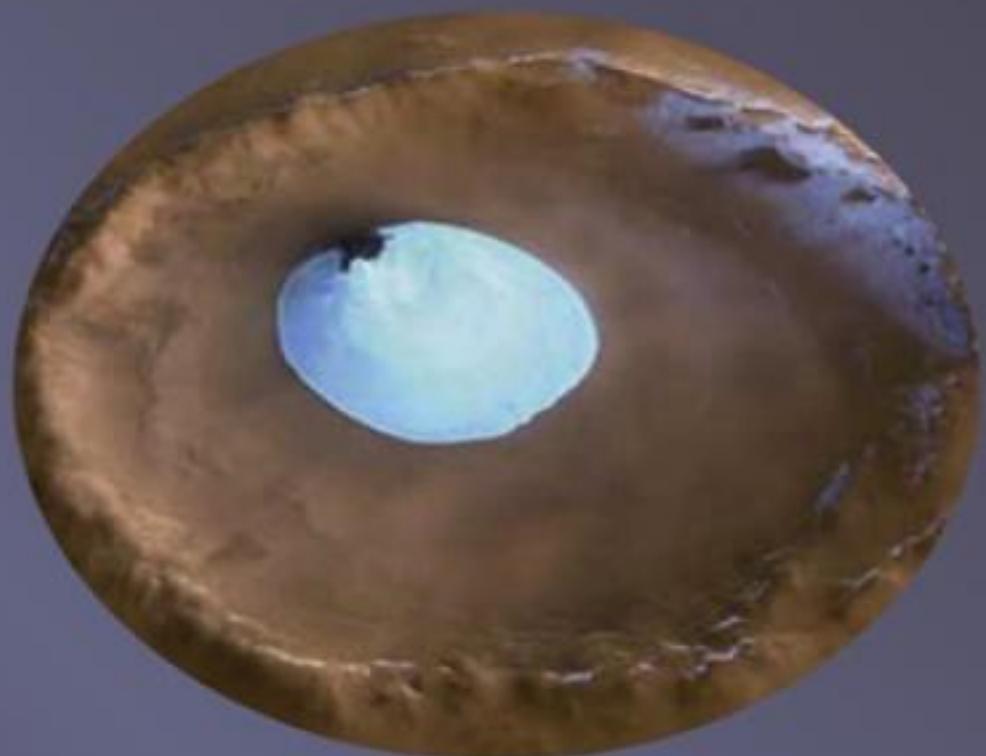




VOLUME 10

TREATISE ON GEOPHYSICS

PLANETS AND MOONS



Volume Editor **Tilman Spohn**

Editor-in-Chief **Gerald Schubert**

Editor-in-Chief

Professor Gerald Schubert

*Department of Earth and Space Sciences and Institute of Geophysics and Planetary Physics,
University of California Los Angeles, Los Angeles, CA, USA*

Volume Editors

Volume 1 Seismology and the Structure of the Earth

Dr. Barbara Romanowicz

University of California at Berkeley, CA, USA

Dr. Adam Dziewonski

Harvard University, Cambridge, MA, USA

Volume 2 Mineral Physics

Dr. G. David Price

University College London, London, UK

Volume 3 Geodesy

Dr. Tom Herring

Massachusetts Institute of Technology, Cambridge, MA, USA

Volume 4 Earthquake Seismology

Dr. Hiroo Kanamori

California Institute of Technology, Pasadena, CA, USA

Volume 5 Geomagnetism

Dr. Masaru Kono

Tokyo Institute of Technology, Tokyo, Japan

Volume 6 Crust and Lithosphere Dynamics

Professor Anthony B. Watts

University of Oxford, Oxford, UK

Volume 7 Mantle Dynamics

Dr. David Bercovici

Yale University, New Haven, CT, USA

Volume 8 Core Dynamics

Dr. Peter Olson

Johns Hopkins University, Baltimore, MD, USA

Volume 9 Evolution of the Earth

Dr. David Stevenson

California Institute of Technology, Pasadena, CA, USA

Volume 10 Planets and Moons

Dr. Tilman Spohn

Deutsches Zentrum für Luft- und Raumfahrt, Berlin, Germany

Preface

Geophysics is the physics of the Earth, the science that studies the Earth by measuring the physical consequences of its presence and activity. It is a science of extraordinary breadth, requiring 10 volumes of this treatise for its description. Only a treatise can present a science with the breadth of geophysics if, in addition to completeness of the subject matter, it is intended to discuss the material in great depth. Thus, while there are many books on geophysics dealing with its many subdivisions, a single book cannot give more than an introductory flavor of each topic. At the other extreme, a single book can cover one aspect of geophysics in great detail, as is done in each of the volumes of this treatise, but the treatise has the unique advantage of having been designed as an integrated series, an important feature of an interdisciplinary science such as geophysics. From the outset, the treatise was planned to cover each area of geophysics from the basics to the cutting edge so that the beginning student could learn the subject and the advanced researcher could have an up-to-date and thorough exposition of the state of the field. The planning of the contents of each volume was carried out with the active participation of the editors of all the volumes to insure that each subject area of the treatise benefited from the multitude of connections to other areas.

Geophysics includes the study of the Earth's fluid envelope and its near-space environment. However, in this treatise, the subject has been narrowed to the solid Earth. The *Treatise on Geophysics* discusses the atmosphere, ocean, and plasmasphere of the Earth only in connection with how these parts of the Earth affect the solid planet. While the realm of geophysics has here been narrowed to the solid Earth, it is broadened to include other planets of our solar system and the planets of other stars. Accordingly, the treatise includes a volume on the planets, although that volume deals mostly with the terrestrial planets of our own solar system. The gas and ice giant planets of the outer solar system and similar extra-solar planets are discussed in only one chapter of the treatise. Even the *Treatise on Geophysics* must be circumscribed to some extent. One could envision a future treatise on Planetary and Space Physics or a treatise on Atmospheric and Oceanic Physics.

Geophysics is fundamentally an interdisciplinary endeavor, built on the foundations of physics, mathematics, geology, astronomy, and other disciplines. Its roots therefore go far back in history, but the science has blossomed only in the last century with the explosive increase in our ability to measure the properties of the Earth and the processes going on inside the Earth and on and above its surface. The technological advances of the last century in laboratory and field instrumentation, computing, and satellite-based remote sensing are largely responsible for the explosive growth of geophysics. In addition to the enhanced ability to make crucial measurements and collect and analyze enormous amounts of data, progress in geophysics was facilitated by the acceptance of the paradigm of plate tectonics and mantle convection in the 1960s. This new view of how the Earth works enabled an understanding of earthquakes, volcanoes, mountain building, indeed all of geology, at a fundamental level. The exploration of the planets and moons of our solar system, beginning with the Apollo missions to the Moon, has invigorated geophysics and further extended its purview beyond the Earth. Today geophysics is a vital and thriving enterprise involving many thousands of scientists throughout the world. The interdisciplinarity and global nature of geophysics identifies it as one of the great unifying endeavors of humanity.

The keys to the success of an enterprise such as the *Treatise on Geophysics* are the editors of the individual volumes and the authors who have contributed chapters. The editors are leaders in their fields of expertise, as distinguished a group of geophysicists as could be assembled on the planet. They know well the topics that had to be covered to achieve the breadth and depth required by the treatise, and they know who were the best of

their colleagues to write on each subject. The list of chapter authors is an impressive one, consisting of geophysicists who have made major contributions to their fields of study. The quality and coverage achieved by this group of editors and authors has insured that the treatise will be the definitive major reference work and textbook in geophysics.

Each volume of the treatise begins with an ‘Overview’ chapter by the volume editor. The Overviews provide the editors’ perspectives of their fields, views of the past, present, and future. They also summarize the contents of their volumes and discuss important topics not addressed elsewhere in the chapters. The Overview chapters are excellent introductions to their volumes and should not be missed in the rush to read a particular chapter. The title and editors of the 10 volumes of the treatise are:

Volume 1: Seismology and Structure of the Earth

Barbara Romanowicz
University of California, Berkeley, CA, USA
Adam Dziewonski
Harvard University, Cambridge, MA, USA

Volume 2: Mineral Physics

G. David Price
University College London, UK

Volume 3: Geodesy

Thomas Herring
Massachusetts Institute of Technology, Cambridge, MA, USA

Volume 4: Earthquake Seismology

Hiroo Kanamori
California Institute of Technology, Pasadena, CA, USA

Volume 5: Geomagnetism

Masaru Kono
Okayama University, Misasa, Japan

Volume 6: Crustal and Lithosphere Dynamics

Anthony B. Watts
University of Oxford, Oxford, UK

Volume 7: Mantle Dynamics

David Bercovici
Yale University, New Haven, CT, USA

Volume 8: Core Dynamics

Peter Olson
Johns Hopkins University, Baltimore, MD, USA

Volume 9: Evolution of the Earth

David Stevenson
California Institute of Technology, Pasadena, CA, USA

Volume 10: Planets and Moons

Tilman Spohn
Deutsches Zentrum für Luft-und Raumfahrt, GER

In addition, an eleventh volume of the treatise provides a comprehensive index.

The *Treatise on Geophysics* has the advantage of a role model to emulate, the highly successful *Treatise on Geochemistry*. Indeed, the name *Treatise on Geophysics* was decided on by the editors in analogy with the geochemistry compendium. The *Concise Oxford English Dictionary* defines treatise as “a written work dealing formally and systematically with a subject.” Treatise aptly describes both the geochemistry and geophysics collections.

The *Treatise on Geophysics* was initially promoted by Casper van Dijk (Publisher at Elsevier) who persuaded the Editor-in-Chief to take on the project. Initial meetings between the two defined the scope of the treatise and led to invitations to the editors of the individual volumes to participate. Once the editors were on board, the details of the volume contents were decided and the invitations to individual chapter authors were issued. There followed a period of hard work by the editors and authors to bring the treatise to completion. Thanks are due to a number of members of the Elsevier team, Brian Ronan (Developmental Editor), Tirza Van Daalen (Books Publisher), Zoe Kruze (Senior Development Editor), Gareth Steed (Production Project Manager), and Kate Newell (Editorial Assistant).

G. Schubert

Editor-in-Chief

10.01 Overview

T. Spohn, DLR Institut für Planetenforschung, Berlin, Germany

© 2007 Elsevier B.V. All rights reserved.

10.01.1	Introduction	1
10.01.2	Our Planetary System	2
10.01.3	Planetary Missions	5
10.01.4	Planet and Satellite Orbits and Rotation States	9
10.01.5	Composition and Interior Structure of Planets	10
10.01.6	Surfaces and Atmospheres	13
10.01.7	Energy Balance and Evolution	17
10.01.8	Magnetic Fields and Field Generation	20
10.01.9	Origin of the Solar System	23
10.01.10	Concluding Remarks	24
	References	25

10.01.1 Introduction

Humanity has always been fascinated with the wandering stars in the sky, the planets. Ancient astrologists have observed and used the paths of the planets in the sky to time the seasons and to predict the future. Observations of the planets helped J. Kepler to formulate his laws of planetary motion and revolutionize the perception of the world. With the advent of the space age, the planets have been transferred from bright spots in the sky to worlds of their own right that can be explored, in part by using the *in situ* and remote-sensing tools of the geosciences. The terrestrial planets are of particular interest to the geoscientist because comparison with our own planet allows a better understanding of our home, the Earth. Venus offers an example of a runaway greenhouse that has resulted in what we would call a hellish place. With temperatures of around 450°C and a corrosive atmosphere that is also optically nontransparent, Venus poses enormous difficulties to spacecraft exploration. Mars is a much friendlier planet to explore but a planet where greenhouse effects and atmospheric loss processes have resulted in a cold and dusty desert. But aside from considerations of the usefulness of space exploration in terms of understanding Earth, the interested mind can visit astounding and puzzling places. There is the dynamic atmosphere of Jupiter with a giant thunderstorm that has been raging for centuries. There is Saturn with its majestic rings and there are Uranus and Neptune with complicated magnetic fields. These giant planets have moons that are similarly astounding. There is the

volcanic satellite of Jupiter, Io that surpasses the Earth, and any other terrestrial planet in volcanic activity and surface heat flow. This activity is powered by tides that twist the satellite such that its interior partially melts. A much smaller moon of Saturn, Enceladus, also has geysers that could be powered by tidal heating. Its volcanic activity releases water vapor not lava. There is another moon of Saturn, Titan, that hides its surface underneath a layer of photochemical smog in a thick nitrogen atmosphere and there are moons of similar sizes that lack any comparable atmosphere. Miranda, satellite of Uranus, appears as if it has been ripped apart and reassembled. Triton, a satellite of Neptune, has geysers of nitrogen powered by solar irradiation. Magnetic field data suggest that icy moons orbiting the giant planets may have oceans underneath thick ice covers. These oceans can, at least in principle, harbor or have harbored life. Moreover, there are asteroids with moons and comets that may still hold the clues to how the solar system and life on Earth formed.

This volume of the *Treatise on Geophysics* discusses fundamental aspects of the science of the planets. It is focused on geophysical properties of the Earth-like planets and moons, those bodies that consist largely of rock, iron, and water, and the processes occurring in their interiors and on their surfaces. But it goes further by discussing the giant planets and their satellites as well. The better part of the volume is dedicated to the interior structure and evolution of the terrestrial planets and to their physical properties such as gravity and magnetic fields, rotation and surface-atmosphere interactions. What is the planetological context of life?

An attempt at answering this question is the subject of a dedicated chapter. In the second half we turn to the outer fringes of the solar system, to the giant planets, their satellites, and to Pluto and the Kuiper Belt. Because exploration by spacecraft is so fundamental to the geophysics of the planets, this volume has two chapters that discuss the orbital dynamics – the routes to the planets – and the instruments typically flown on planetary missions. Limits in space keep us mostly from discussing the small bodies (asteroids, comets, and dust) as well as from discussing to a larger extent planets in other solar systems. The latter bodies are too far away to be explored *in situ*, a prerequisite for geophysical methods to be applied. Exploration of these bodies remains for the time being in the realm of astronomy.

10.01.2 Our Planetary System

The solar system contains a myriad of bodies ranging in size from the Sun to minuscule dust particles. The Encyclopedias of Planetary Sciences (Shirley and Fairbridge, 1997) and the Solar System (Weissmann *et al.*, 1999) and the Planetary Companion (Lodders and Fegley, 1998) are useful sources of information about the solar system. Valuable collections of planetary data can also be found on Solar System Dynamics and National Space Science Data Center websites. A collection of images of planetary surfaces can be found in the NASA photojournal.

The eight planets Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune are shown in **Figure 1** together with Pluto the most prominent member of a new class of objects, the dwarf planets. The International Astronomical Union (IAU) introduced the class of dwarf planets in 2006. These are intermediate in size between the terrestrial planets and small bodies such as cometary nuclei and most asteroids. The IAU has so far identified the following three celestial bodies as dwarf planets – Pluto, the asteroid Ceres, and the Trans-Neptunian Object (TNO) Eris, a.k.a. 2003UB313. Other candidate bodies such as asteroid Vesta and the TNO Sedna are under consideration. A planet of our solar system according to the IAU resolution 5 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 according to the resolution satisfies (1) and (2) but has NOT cleared its neighborhood and (4) is not a satellite. See also Basri and Brown (2006) for a discussion of the term ‘planet’.

The Sun, a middle-aged main sequence star, contains 98.8% of the mass of the solar system but only 0.5% of its angular momentum. The next smaller body, Jupiter, still 300 times more massive than Earth (see **Table 1**), contains more than 60% of the mass of the rest. Jupiter is the biggest of the ‘giant planets’, a group of gaseous planets that constitute a major subgroup of the solar system (*see* Chapter 10.13).

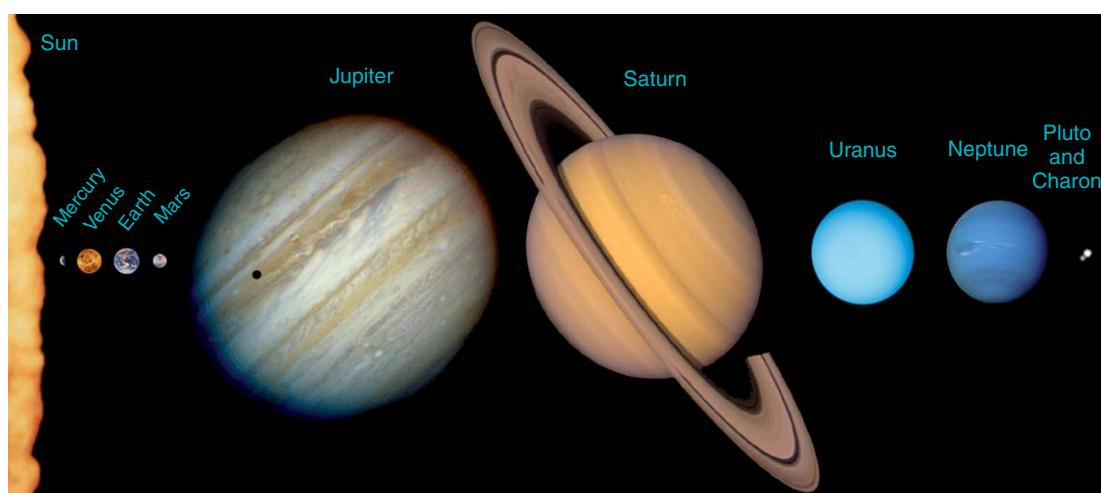


Figure 1 The eight planets Mercury, Venus, Earth, Mars, Jupiter, Saturn, Uranus, and Neptune and the dwarf planet Pluto are shown in this compilation of NASA images with their correct relative sizes and ordered according to their distance from the Sun. Mercury is barely visible at left close to the arc of the surface of the Sun and Pluto is barely visible at the outer right. The dark spot on Jupiter is the shadow of Io, one of its major satellites. © C.J. Hamilton.

Table 1a Properties of the planets

	Terrestrial or Earth-like Planets				Giant Planets			
	Mercury	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune
Radius (km)	2438.	6052.	6371.	3390.	71492.	60268.	24973.	24764.
Mass (10^{24} kg)	0.3302	4.869	5.974	0.6419	1899.	568.46	86.63	102.4
Density (10^3 kg m $^{-3}$)	5.430	5.243	5.515	3.934	1.326	0.6873	1.318	1.638
Uncompressed Density (kg m $^{-3}$)	5.3	4.0	4.05	3.75	0.1	0.1	0.3	0.3
Rotational Period (d a)	58.65	243.0 ^b	0.9973	1.026	0.4135	0.4440	0.7183 ^b	0.6713
Inclination of rotation axis (°)	0.5	177.4	23.45	25.19	3.12	26.73	97.86	29.56
Orbital distance (AU c)	0.3871	0.7233	1.000	1.524	5.203	9.572	19.19	30.07
Orbital Period (a d)	0.2410	0.6156	1.001	1.882	11.87	29.39	84.16	165.0
Magnetic Moment (10^{-4} T \times radius 3)	3×10^{-3}	$<3 \times 10^{-4}$	0.61	$<6 \times 10^{-4}$	4.3	0.21	0.23	0.133
Effective Surface Temperature (K)	445	325	277	225	123	90	63	50
Specific heat flow or luminosity (pW kg $^{-1}$)	?	?	7	?	176	152	4	67
Known satellites	0	0	1	2	65	56	17	8

Table 1b Properties of dwarf planets

	Ceres	Pluto	Eris
Radius (km)	476.2	1195	1200 ± 50
Mass (10^{24} kg)	9.46×10^{-4}	0.01305	?
Density (10^3 kg m $^{-3}$)	2.08	2.03	?
Rotational period (d a)	0.3781	-6.387 ^b	?
Orbital distance (semimajor axis) (AU c)	2.766	39.48	67.73
Orbital period (a d)	4.6033	248.1	557.8
Orbital eccentricity	0.07976	0.249	0.4400
Orbital inclination	10.59	17.14	44.16
Effective surface temperature (K)	167	44	~30
Known satellites	0	1 (Charon)	1 (Champollion)

Table 1c Properties of major satellites

	Moon			Ganymede		Titan	Triton
	Earth	Io	Europa	Jupiter	Callisto	Saturn	Neptune
Primary							
Radius (km)	1737.	1821.	1560.	2634.	2400.	2575.	1353.
Mass (10^{20} kg)	734.9	891.8	479.1	1482.	1077.	1346.	214.7
Density (10^3 kg m $^{-3}$)	3.344	3.53	3.02	1.94	1.85	1.881	2.054
Specific heat flow (pW kg $^{-1}$)	8	890	150(?)	?	?	?	?
Orbital period (d)	27.32	1.769	3.551	7.155	16.69	15.95	5.877 ^b

^aA day (d) is equivalent to 24 h.^bThe motion (rotation, revolution) is retrograde.^cAU is one astronomical unit or 149.6 million kilometers.^dA year (a) is equivalent to 365 days (d). The uncompressed density is model dependent and cannot be given with more than 2 to 3 significant digits.

Source: The data have been taken from the compilation of Lodders and Fegley, The Planetary Companion, Oxford University Press, 1998, from Spohn, Planetologie, in Bergmann-Schäfer, Lehrbuch der Experimentalphysik, Vol. 7, (ed. W. Raith), W. De Gruyter, Berlin, pp. 427–525. and from <http://nssdc.gsfc.nasa.gov> and <http://ssd.nasa.gov>. For the sake of space, we give four significant digits even for numbers that are known more accurately.

4 Overview

Among the giant planets are, in addition to Jupiter, Saturn, Uranus, and Neptune. The latter two are sometimes called the subgiants or the ice giants because they are notably smaller than Saturn and Jupiter and because they mostly consist of water, methane, and ammonia, components often collectively called the planetary ices (see below). Earth, the biggest member of the other major subgroup of family members, the ‘terrestrial planets’, is the only planet on which we know that life had a chance to develop. Among the members of this group are Mercury, the innermost planet, Venus, Earth’s twin with respect to size and mass, and Mars. The latter planet has the best chance of having developed some primitive forms of life, which makes it the prime target of present-day space missions. The terrestrial planets together have about 0.005% of the mass of the solar system. **Table 1** collects some data of general interest on the planets and some major moons.

The rotational periods of the major satellites are equal to their orbital periods. This is known to be the case for all satellites listed except for Titan for which the rotational period is not (yet) known.

Comparative planetology is the science of studying the planets by comparing and finding general properties and common lines of evolution as well as features that are specific. It uses the methods of the natural sciences and is strongly interdisciplinary. Comparative planetology is not restricted to the planets *sensu strictu* but also considers the major moons of the planets such as Earth’s Moon, the major satellites of Jupiter (Io, Europa, Ganymede, and Callisto) (**Figure 2**), the major Saturnian satellite Titan and, finally Triton, Neptune’s major satellite. Other bodies of interest include the yet largely unexplored members of the asteroid belt, the asteroids, the comets, and the Kuiper Belt objects. The asteroid belt is found between Mars and Jupiter and its members are speculated to be the parent bodies of most meteorites (stones from space

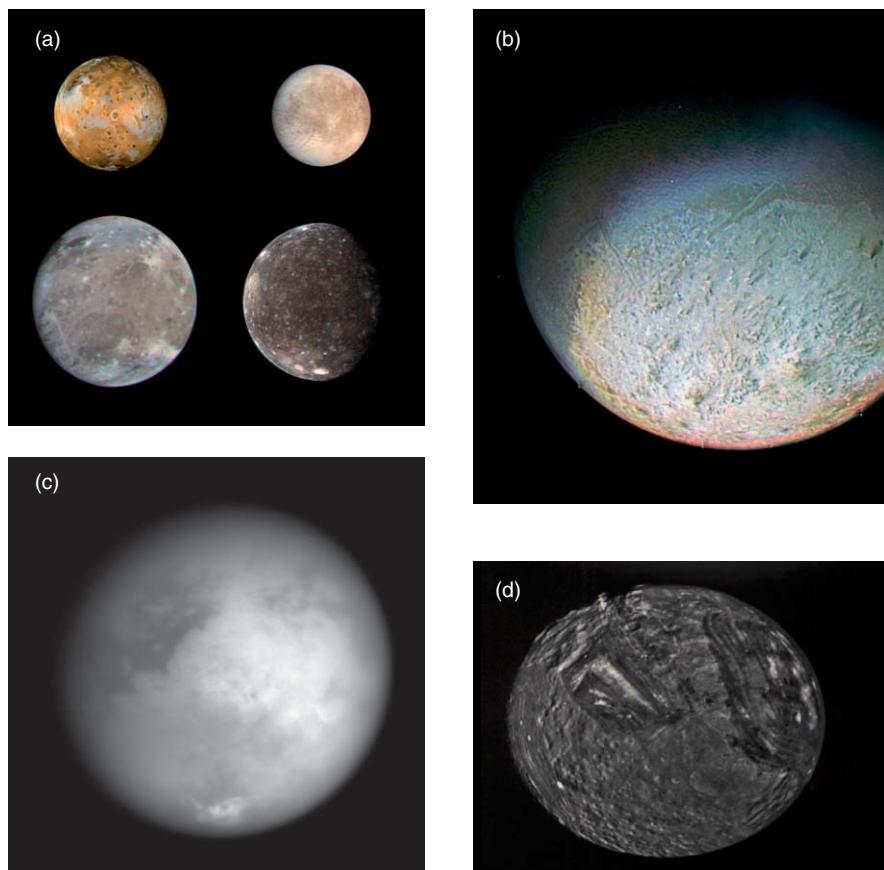


Figure 2 The four Galilean satellites of Jupiter Io, Europa, Ganymede, and Callisto, the Saturnian satellite Titan, the Uranian satellite Miranda, and the Neptunian satellite Triton are shown in this compilation of spacecraft images from top left to bottom right (not to scale). © NASA/JPL.

found on Earth's surface compare articles in the *Treatise of Geochemistry*). The Kuiper Belt stretches between 30 AU (roughly the orbit of Neptune to at least 50 AU and contains Pluto, Sedna, and Eris. The asteroids are believed to be the source of most meteorites and contain rich information about the formation of the solar system, its evolution, and even the evolution of matter before the formation of the solar system. The Kuiper Belt objects are basically unexplored. Pluto, the most prominent member, and the Kuiper Belt are the target of the New Horizons mission launched in 2006 (see Chapter 10.16).

10.01.3 Planetary Missions

The results of planetary sciences are made possible at their present levels only through the exploration of the solar system with space missions (see, e.g., Friedman and Kraemer, 1999, and Chapters 10.17 and 10.18). All planets and major satellites have been visited at least by a flyby of a robot spacecraft. This is not true for the dwarf planets but missions have been launched to Pluto (New Horizons) or will soon (2007) be launched to Ceres (Dawn). Some, such as Venus, the Moon, Mars, Jupiter, and Saturn have been explored by orbiters and by landers. The Moon has even been visited by astronauts. It can be said without any doubt that space exploration has turned dim discs in the sky observable with telescopes into worlds of their own rights waiting for further, perhaps eventually human exploration.

Space exploration began with missions to the Moon in the 1960s, characterized by the race to the Moon between the US and the then Soviet Union. This race culminated in the landing of men on the Moon. For planetary sciences the significance of the Apollo Programme lies mainly with the first sample return, the first seismic exploration of a planetary interior, the first heat flow measurement, and the first geologic field trip on another planet. Although Apollo has not finally solved the puzzle of the origin of the Moon – some may argue that it can never be solved once and for all – it has provided a wealth of data and has made the Moon one of the best-known planetary body other than the Earth. The planet most frequently visited by spacecraft aside from the Moon is Mars. Highlights of Martian exploration include the Viking missions (the first landings on Mars), the Pathfinder mission (the first vehicle introducing mobility), Mars Express (the first European planetary mission *sensu strictu*), and the Mars Exploration Rovers (MER for short) (Figure 3). From a geophysics point of view, Mars

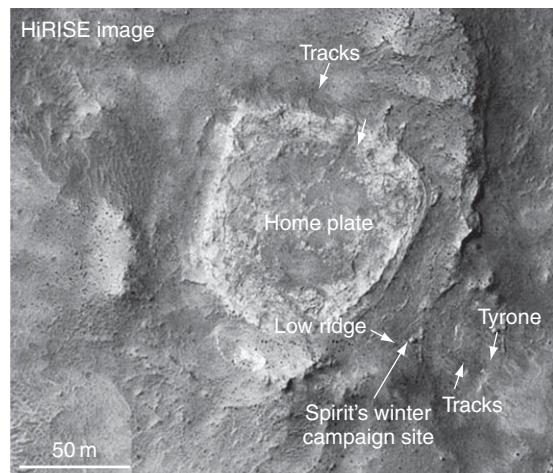


Figure 3 Traces of the MER rovers imaged by the ultimate resolution camera on board the Mars Reconnaissance Orbiter. © NASA/JPL.



Figure 4 This image has been interpreted to show a frozen over lake on Mars comparable in size to the North Sea on Earth. © ESA/DLR/FU Berlin (G. Neukum).

Global Surveyor which returned highly accurate maps of topography and gravity (Smith *et al.*, 1999) and detected the remnant magnetization of old crustal units (Acuna *et al.*, 1999) has been a remarkable mission. MER is doing the first geologic field trip on Mars and detected on their now tens-of-kilometers-long journey minerals that formed under humid conditions (Squyres and Knoll (2005) and references therein). The Mars Express mission detected methane on Mars, numerous morphological units that required water to form (see Chapter 10.11) and showed that volcanic activity continued albeit at a low rate until the recent past (Neukum *et al.*, 2004) (Figure 4). Future Mars missions will bring network stations to Mars with seismometers and heat flow probes and samples back to Earth.

6 Overview

An important driver for Mars research is its potential habitability and the possible detection of life, extinct or present. The Viking landers in the 1970s already attempted to find signatures of bioactivity but failed to do so. Beagle 2 onboard the European Mars Express Mission was built to search for bioactivity and would have been the first European lander on another planet had it not failed. The next attempt to directly detect biosignatures on Mars will be the ExoMars rover scheduled for launch in 2013. ExoMars may also be the mission to land the first geophysics package on Mars with a seismometer, heat flow probe, and atmospheric science package. The first seismometer was brought to Mars onboard Viking but failed to return useful data because of bad coupling to the surface (*see* Chapter 10.03).

Venus has been the target of very successful Soviet missions that included eight Landers (Venera 7–14). These landers survived in the highly corrosive atmosphere of Venus (see below) only for a few hours but transmitted color photos from the surface which shape our image of the Venusian surface. Further highlights of Venus exploration have been the Pioneer Venus and Magellan missions which provided extensive data on the atmosphere of Venus and the first systematic mapping of the surface with radar. At the time of this writing Venus Express, the first European mission to Venus and Mars Express' sister spacecraft, is orbiting the planet. The

first results show images at the South-Polar atmosphere vertex and of the surface in the infrared part of the electromagnetic spectrum. Temperature in **Figure 5** is inversely correlated with topographic height.

Since the atmosphere of Venus is optically thick, cameras are of little use at this planet for surface exploration but are used to study atmosphere circulation patterns. Mercury, the innermost planet, has been visited only through two flybys of the American Mariner 10 mission targeted for Mars. Mercury is the target of the Messenger mission launched in 2004 and the ambitious European BepiColombo mission scheduled for launch in 2013.

The outer solar system was first explored by the Pioneers that flew by Jupiter (Pioneer 10 and 11) and Saturn (Pioneer 11) and subsequently by the very successful Voyager I and II missions that flew by all four major planets of the outer solar system. In addition to exploring the planets, these spacecraft discovered rings at Jupiter and Uranus and explored the satellite systems and found Io to be the most volcanically active body in the solar system. The Voyagers for the first time allowed a detailed view of the rich phenomenology of icy satellite surfaces. After the Pioneer and Voyager flybys orbiters were the next logical step in outer solar system exploration. The first mission in this class was Galileo which orbited Jupiter and plunged into Jupiter's crushing

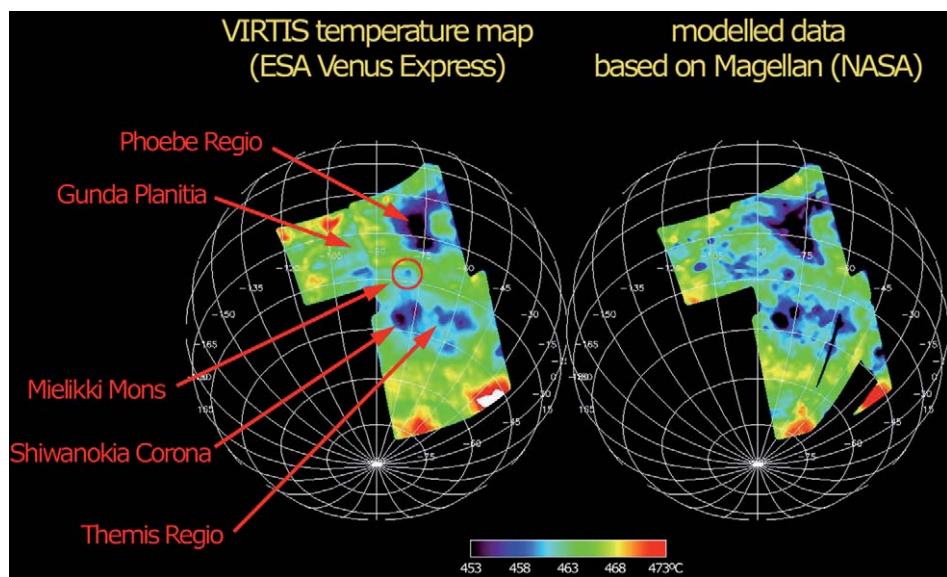


Figure 5 Map of half of the southern hemisphere of Venus in the infrared taken with the VIRTIS infrared spectrometer on Venus Express. © ESA.

atmosphere in September 2003. The spacecraft was deliberately destroyed to protect the Jovian system, in particular Europa, from being polluted. Galileo discovered the magnetic field of Ganymede and evidence for oceans underneath the ice crusts of the Galilean moons Europa, Ganymede, and Callisto. On its route to Jupiter Galileo discovered for the first time a moon (Dactyl) that is orbiting an asteroid (Ida) and observed comet Shoemaker-Levy crashing into Jupiter. Galileo imaged the Jovian satellites at much improved resolution and discovered several new ones. Comparison with Voyager images of volcanic features on Io showed significant modifications that had occurred in the roughly 20 years between the missions. An instrumented descent probe released from the Galileo orbiter entered the Jovian atmosphere in December 1995 and provided the first *in situ* measurements of the state and the chemistry of a giant planet atmosphere shroud. The results of the Galileo mission are discussed in Bagenal *et al.* (2004).

At the time of this writing Cassini is orbiting Saturn. Cassini is expected to top the success of Galileo as an outer solar system exploration mission through its more advanced payload capability. The Huygens probe supplied by the European Space Agency ESA descended through the atmosphere of Saturn's biggest moon Titan and successfully landed on the surface on 14 January 2005 to transmit the first images (color) from Titan's surface (Figure 6). Titan's atmosphere is optically thick like that of Venus. Huygens landed on a solid surface strewn with ice boulders. On its descent it imaged river beds in which hydrocarbons are flowing or have flowed in the past and a possible shoreline (Lebreton *et al.*, 2005). Near-infrared data from the Cassini mission have been interpreted as showing a cryovolcano on Titan (Sotin *et al.*, 2005). Another highlight of the mission is the detection of active ice volcanism on the small satellite Enceladus (Porco *et al.*, 2006) the diameter of which is only 500 km and a rift structure circling Iapetus along its equator (Porco *et al.*, 2005) (Figure 7). Cassini also returned gravity data that allowed models of the interiors of some of the smaller satellites (*see* Chapter 10.15). Cassini will continue to explore Saturn, the rings, and its moons in a 4-year orbiting journey through the system. Among the most spectacular planned missions for the outer solar system is a lander mission for Europa. The latter mission is driven by the perception that an ocean on Europa may be



Figure 6 First color image from Titan's surface. This image was processed from images returned on 14 January 2005, by ESA's Huygens probe during its successful descent to land on Titan. The rock-like objects just below the middle of the image are about 10 cm across. The surface consists of a mixture of water and hydrocarbon ice. There is also evidence of erosion at the base of these objects, indicating possible fluvial activity. © ESA/NASA/JPL/University of Arizona.

covered by a thin ice shell that can be drilled and by the possibility that life developed in this ocean.

Asteroids and comets are believed to be remnants of accretion and to consist of largely primitive matter. Although this perception is questionable to some extent – comet nuclei that have repeatedly flown by the Sun may have experienced thermal metamorphism – some asteroids show signs of endogenic activity

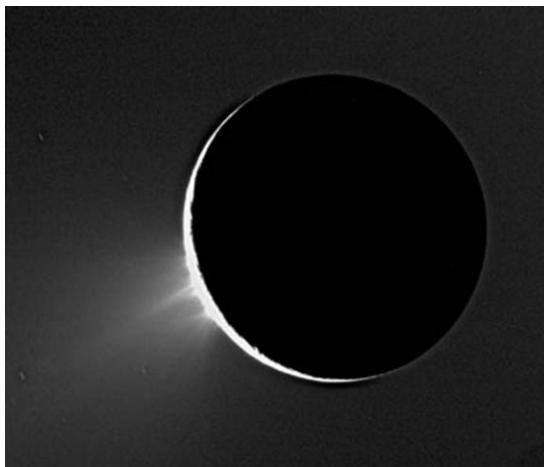


Figure 7 Geysers emanating from the southern hemisphere of Enceladus. © NASA/JPL.

and may be differentiated (compare Volume 1 of the *Treatise in Geochemistry*); asteroids and comets are minor solar system bodies of great interest. The Dawn mission will explore the two largest asteroids Vesta and Ceres. The first is believed to be a differentiated body while the latter is believed to be undifferentiated and mostly primitive. Comet nucleus Churyumov-Gerasimenko is the target of the ESA Rosetta mission (launched in 2004). After a 10-year cruise the Rosetta orbiter will begin to orbit the nucleus while the lander Philae will dock onto the nucleus and explore it *in situ*. Both the orbiter and Philae will take data as the comet approaches the Sun.

Planetary exploration has taken and is still following a stepped approach by which flybys are followed by orbiter missions. These are followed by lander missions including networks of landing stations and rovers. Sample return to the Earth is then foreseen, followed by human exploration. The latter step is often met with skepticism because of its prohibitive cost and, except for the Moon, is still facing huge problems of a technical and medical nature.

The payloads of orbiters and flyby spacecraft are specifically designed for their science goals (see Chapter 10.18). This is even more true today after some major space mission failures have resulted in smaller, more dedicated missions. It is still possible, however, to characterize generic payload elements that are typically found on these spacecraft. Almost every mission has at least one camera, often two for wide-angle and for high resolution. Early missions had panchromatic cameras but recent missions have color capabilities although not necessarily at the

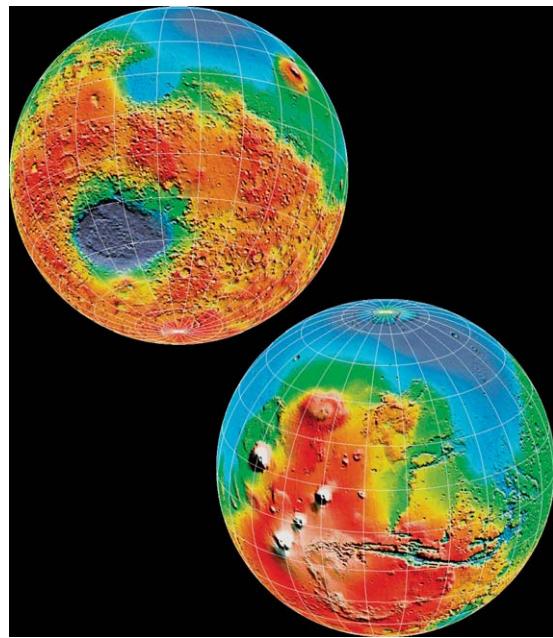


Figure 8 Global map of Mars taken by the Mars Observer Laser Altimeter MOLA onboard the Mars Global Surveyor mission. © NASA/JPL.

same resolution. Stereo is a useful feature that can be found on missions like Mars Express. Stereo allows three-dimensional (3-D) imaging and enables development of digital terrain models. Digital terrain modeling is helped by laser altimetry as the successful Mars Observer Laser Altimeter (MOLA) on Mars Global Surveyor has so convincingly demonstrated (**Figure 8**). MOLA has made laser altimeters almost a must on a modern exploration spacecraft. Laser altimeters provide a nice example of how technology helps to save the most precious resources on spacecraft, power and mass. While MOLA had a 26-kg mass its successors, the Laser Altimeters MLA and BELA, built for the Messenger and BepiColombo missions to Mercury, are only about half as massive. Next-generation laser altimeters will use single-photon techniques that will further save mass and, in addition, significantly cut down on power. As geodesy tools providing the topography of a planet the two are complemented by radio science instrumentation to measure the gravity field. In its simplest form, radio science instrumentation uses the communication devices of the spacecraft and a suitably accurate clock to measure Doppler shifts of radio signals traveling to and from the spacecraft. More advanced systems use their own high-frequency radio channels and accelerometers to measure the

changes in spacecraft motion. If the surface is hidden underneath an optically thick atmosphere radar can be used to map the surface as has been successfully done at Venus by the Magellan mission. Radar is further used to map the surface roughness and to penetrate of the near-surface layers, for instance to look for ground ice as is done by Mars Express. Other microwave sounders have been proposed that use different wavelengths and have differing depths of penetration.

While cameras, laser altimeters, and radio science instrumentation are indispensable instruments for planetary geology and geodesy, spectrometers are the tools of cosmochemistry and mineralogy. These instruments typically analyze particles (photons, phonons, neutrals, ions) emitted from surfaces and atmospheres by the incident solar radiation. Visible, near- and thermal infrared spectrometers are used to characterize the mineralogy of surface rock (silicates, carbonates, sulfates, etc.) and to search for water ice. Mapping infrared spectrometers such as Themis (0.4–15 μm) on Mars Odyssey and Omega (0.5–5.2 μm) on Mars Express allow a mapping of minerals on the surface. The Planetary Fourier Spectrometer on Mars Express works in a similar but broader spectral range (1.2–45 μm) but has a much higher spectral resolution at the expense of spatial resolution. Ultraviolet spectrometers measure neutral metals such as Al, Na, and S, ozone and OH-radicals. Gamma-ray spectrometers can measure radionuclides (K, Th, and U), major rock-forming elements (O, Mg, Al, Si, Ca, Ti, and Fe), hydrogen (from water), and carbon, depending on the energy resolution and the elemental concentration in the surface. X-ray spectrometers are suitable to detect the major elements Mg, Al, Si, Ca, and Fe.

The particles and electromagnetic fields surrounding planets and other solar system objects are measured with magnetometers and plasma detectors. Dust analyzers, finally, collect dust particles and analyze the directions and strengths of dust flows in space.

Landing missions carry cameras and have been equipped with spectrometers and other sensors to measure composition and mineralogy. Mobility is an important issue since exploration of just one landing site may lead to biased results. Recent examples are the two MER rovers that carry a panoramic stereo camera, an alpha-particle X-ray spectrometer for elemental composition, a Mössbauer spectrometer for the identification of iron-bearing minerals, a miniaturized thermal infrared

spectrometer for mineralogy, and a rock-abrasion tool to allow the measurement of ‘fresh’ rock (Squyres *et al.*, 2003). Network landers have not been flown yet but are generally agreed to be important (cf. Chapter 10.03). These landers will operate the same payload simultaneously at several locations. Applications are seismology to explore the interior of planets, meteorology to explore planetary atmospheres, electromagnetic induction studies, and planetary geodesy. But single landers still have important applications such as the ill-fated Beagle Lander on ESA’s Mars Express that was geared to search for traces of extinct and extant life on Mars and that failed upon landing in late 2003. Although lander missions carry a significantly higher risk compared to orbiter missions, it is generally agreed that these will be indispensable tools of future space exploration. Thus NASA is planning to have the Mars Science Lab rovers in 2009 and ESA is preparing for the ExoMars mission with the Pasteur rover to be launched in 2013.

10.01.4 Planet and Satellite Orbits and Rotation States

There are some interesting commonalities between the orbits and rotation states of the planets and their major satellites. The rotation of the terrestrial planets is discussed in detail in Chapter 10.04. The orbits lie mostly in a plane that is defined by Earth’s orbital plane, termed the ecliptic, and which is close to the equatorial plane of the Sun. The formal definition of the ecliptic plane adopted by the IAU calls for the ecliptic pole to be the mean orbital angular momentum vector of the Earth–Moon barycenter.

The normals to the actual orbital planes have inclinations relative to the ecliptic normal that are small, only a few degrees. The dwarf planets Ceres, Eris, and Pluto differ with orbital inclinations of $\sim 10^\circ$, 44° , and 17° . Small bodies (asteroids and comets) also tend to have larger orbital inclinations. The rotation axes of most planets and major satellites are within a few tens of degrees to the vertical to their orbital planes. Notable exceptions are Uranus, whose rotation axis lies almost in its orbital plane, Venus, whose retrograde rotation can be expressed as an inclination of almost 180° , and Pluto, whose inclination is between those of Uranus and Venus. The reasons for these anomalies are unknown but are sometimes speculated to be attributable to impacts

of planet-sized bodies on the young planets during the late stage of accretion.

The rotation and revolution of most planets and moons are prograde, that is counterclockwise if viewed from above the celestial North Pole. Exceptions are Venus, Uranus, and Pluto whose retrograde rotations can also be described as inclinations of more than 90° of their rotation axes to their orbital plane normals. Another exception is Triton, whose retrograde orbital motion about Neptune is unusual. Triton, like Pluto, is speculated to have originated in the Kuiper belt. It was likely captured into its retrograde orbit by Neptune (see Chapter 10.14). There are many more satellites with retrograde orbits. However, these are typically much smaller than Triton, objects in a size range of only a few hundred kilometers. The known rotation periods of the major satellites and of most satellites in general are equal to their orbital periods. This is termed a 1:1 spin–orbit coupling and is believed to be the result of tidal evolution (see Chapters 10.14 and 10.16). Tidal evolution occurs because planets raise tides on their satellites just as the Moon raises tides on Earth. The gravitational force of the planet then pulls on the tidal bulge of the satellite. This torque will reduce the rotation rate of the satellite – should it be greater than its average orbital angular speed – and the interaction will end with the observed 1:1 coupling. As a result, the satellites present their primaries with mostly the same face at all times.

The planet Mercury is particularly interesting with respect to spin–orbit coupling since it is the only planet that is in such a resonance state with the Sun. However, its coupling is not 1:1, as one may expect, but 3:2 (see Table 1). The reason for this somewhat odd ratio is believed to lie with the unusually large eccentricity of Mercury's orbit that has a value of 0.2 (Peale, 1988; see also Chapter 10.04). Values of less than 0.1 are more typical of planets and major satellites. The dwarf planets and minor bodies have larger eccentricities. The large eccentricity of Mercury's orbit causes significant differences to arise between the constant rotational angular velocity and the orbital angular velocity that varies along the orbit. The two angular velocities could only be exactly equal at all times if the orbit were perfectly circular with zero eccentricity. The orbital velocity on an eccentric orbit increases towards the perihelion (the point of the orbit closest to the Sun) and decreases towards the aphelion (the point farthest from the Sun). Mercury's 3:2 resonance causes the orbital angular velocity to be equal to the rotational

angular velocity at perihelion. This minimizes the tidal torque on Mercury and stabilizes the resonance.

The orbital distances of the planets from the Sun roughly follow a law with the distance of one planet to the Sun being roughly twice the distance to its inner neighbor. This rule is called the 'Titius-Bode law' and works with Jupiter and Mars only if the asteroid belt is counted as a planet. The origin of the law is little understood and it is a condition to be used as a test for accretion models. Regular relations between orbital distances or periods (the latter two are coupled through Kepler's third law of orbital motion) are not unusual in the solar system, however. The most prominent example is the Laplace resonance between the innermost three major satellites of Jupiter: Io, Europa, and Ganymede. A comparison of the orbital periods in Table 1 shows that these are in the ratio of 1:2:4. The origin of the Laplace resonance and its stability is attributed to tidal interactions between Jupiter and its three resonant moons. The orbits expand in the resonance as rotational energy is transferred through tidal interaction from Jupiter to Io and passed on in part to Europa and Ganymede (see, e.g., Yoder and Peale (1981) or Spohn (1997) and references therein). The age of the resonance and how it formed are unknown. From the amplitude of libration it has been concluded that the resonance is relatively young. Thermal-orbital modeling, however, suggests an age of at least 2 billion years. The tidal interaction and, in particular, the dissipation of tidal energy are believed to be the cause of volcanic activity on Io and of a subsurface ocean on Europa (Chapter 10.15). This ocean is completely covered by icebergs and ice shields that may slowly move relative to each other and the ocean may even harbor, or may have harbored, primitive forms of life (see Chapter 10.12). The mathematical theory of tidal evolution is outlined in an appendix to Chapter 10.14.

10.01.5 Composition and Interior Structure of Planets

The chemical composition of the solar system is mainly dominated by the composition of the Sun, which has 98.9% of the mass of all bodies in the system. Although the compositions of the planets are different from that of the Sun they are related to the Sun's composition and reflect varying degrees of depletion in volatile elements (see articles in Volume 1 of the *Treatise on Geochemistry*; Davis

(2004)). The Sun is mainly composed of hydrogen and helium but contains all the other elements found in the solar system in abundances that are characteristic of the so-called ‘solar composition’.

These abundances are obtained from solar spectroscopy and from the analysis of primitive meteorites, the CI chondrites. Hydrogen and helium are dominant for Jupiter and Saturn, which are massive enough to keep these elements in molecular form against their tendency to escape to space. The potential of a planet to keep a specific element is measured by its escape velocity which is proportional to the square root of the ratio between the planet’s mass and its radius and must be compared with the thermal speed of an element which is inversely proportional to its molecular weight. Still, Jupiter and Saturn are depleted in both H and He with respect to the composition of the Sun (*see Chapter 10.13*). In addition to H and He, Jupiter and Saturn contain substantial amounts of water (H_2O), ammonia (NH_3), and methane (CH_4), compounds collectively known as the planetary ices because of their occurrence on the surfaces of the major planets’ icy satellites. In the deep interiors of the giant planets, these compounds are to be found in their supercritical forms for which there is no difference between the gaseous and liquid states. In addition to H, He, and planetary ices, Jupiter and Saturn have central cores of rock and iron.

Models of the interior structure of planets can be constructed from a sufficiently detailed knowledge of their ‘figures’ and their ‘gravity fields’. (The gravity fields of the terrestrial planets are discussed in Chapter 10.05; interior structure models of the terrestrial planets in Chapter 10.02. Chapter 10.04 discusses the rotation of these planets and how the rotation depends on interior structure. The giant planets are discussed in Chapter 10.13 while interior structures of giant planet satellites are discussed in Chapter 10.15.) The construction of interior structure models is a primary task of planetary physics. Rotation deforms the planets from spheres into prolate spheroids. The flattening of both the figure and the gravity fields are dependent upon the variation of density – and thus composition – with depth. Unfortunately, these data do not allow unique models. The most accurate method of exploring interior structure is ‘seismology’ (*see Chapter 10.03*), the study of sound waves that travel through the interiors of planets and of global oscillations. This method is the prime method for exploring the Earth’s interior to provide models of the variation of density with depth and also laterally by a technique that is known as seismic tomography (*see Volume 1*).

Seismometers, the instruments to record the waves, have also been placed on the Moon during the Apollo program and the interpretations of the results have provided important insights, though the results are not completely satisfactory. The acoustic coupling between the surface and the atmosphere offers chances to do seismology from the orbit using cameras, infrared mappers, and ionospheric sounders. This method offers particularly good chances for Venus where environmental conditions on the surface are exceedingly demanding. The method can also be applied, at least in principle, at the giant planets to study free oscillations and constrain interior models with these.

Models of the giant planets (**Figure 9**) suggest that the outer layers of both Jupiter and Saturn are mostly H and He. H is molecular at moderate pressures but becomes metallic at pressures larger than 170 GPa. This pressure is equivalent to depths of about 14 000 km in Jupiter and 27 000 km in Saturn ([Hubbard and Marley, 1989](#); [Guillot, 2005](#); *Chapter 10.13*). The phase transformation occurs at greater depth in Saturn because the pressure increases at a smaller rate in this planet due to its smaller mass. The deeper interiors of both contain the denser ices possibly in a shell above the densest iron/rock core. The cores plus the ice shells are believed to have masses that are quite uncertain between less than 10 and 45 Earth masses and radii of a few Earth radii.

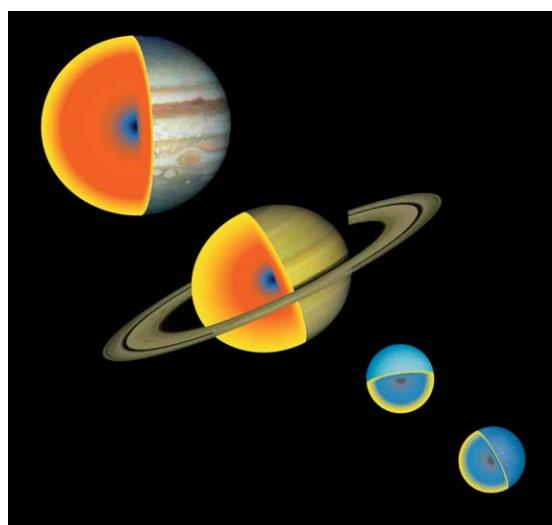


Figure 9 Interior structures of the giant planets. The yellow color indicates the molecular H, He shroud. Orange indicates the metallic H, He layers; blue is the range where the ice components are found and the dark colors indicate the rock/iron cores. Courtesy of T. Guillot.

Uranus and Neptune are further depleted in H and He and consist mainly of ice and the rock/iron component in addition to some H and He. Their structure, as interpretations of the gravity field suggest (Hubbard and Marley, 1989; Chapter 10.13), is not as clearly layered as those of Jupiter and Saturn. Rather, there appears to be a gradual increase of density with depth accompanied by a gradual increase of the abundance of ice at the expense of H and He followed by a gradual increase of the abundance of rock/iron at the expense of ice.

The major satellites of the outer solar system and Pluto also contain substantial amounts of ice, as their densities, around 2000 kg m^{-3} , suggest (see Table 1; Chapter 10.15). Examples to the contrary, with substantially greater densities are Io and Europa the two innermost Galilean satellites of Jupiter. It is tempting to speculate that these two also started their evolution with a substantial abundance of ice. Io may have lost the ice completely and Europa to a large extent as a consequence of heating by impacts or as a consequence of tidal heating. Since Io and Europa are closer to Jupiter than Ganymede and Callisto, both the energy of impactors and the tidal dissipation rate will have been larger and may have caused the loss of water. But it is not clear that tidal heating could have provided the power needed to vaporize or partly vaporize ice layers on the two. However, it is also plausible that the temperature in the Jovian nebula from which the satellites formed was too hot at Io's and Europa's orbital distance to allow the accumulation of substantial ice shells. (The accretion of the satellites of Jupiter is discussed together with the formation of other satellites in Chapter 10.14.) The Galileo mission has returned two-way Doppler ranging data that allow reasonable modeling of their interior structure (Figure 10). (See Chapter 10.15 and references therein for the icy satellites and, e.g., Sohl *et al.* (2002) and Schubert *et al.* (2004) for recent reviews including Io. See also Chapter 10.18 for a brief discussion on radio science data.) Accordingly, Io has an iron-rich core of about half the satellite's radius and 20% of the satellite's mass with a rock shell above. Europa has a (water) ice shell about 150 km thick overlying a silicate rock shell and an iron-rich core. The nonuniqueness of the models allows a wide range of sizes of these reservoirs but the most likely radius values are 500–600 km for the core and 800–1000 km for the rock mantle. It is widely believed that a large part of the ice shell may actually be liquid allowing for an internal ocean. This is suggested by the results of the

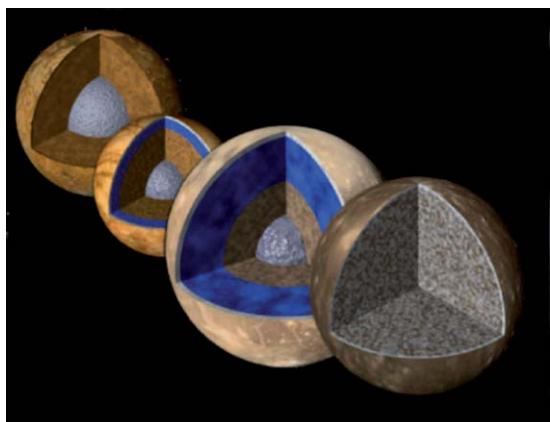


Figure 10 Interior structures of the Galilean satellites of Jupiter, Io, Europa, Ganymede, and Callisto (background to foreground). © NASA/JPL.

Galileo mission magnetic field measurements that indicate a field induced by Europa's motion in the magnetic field of Jupiter. This interpretation of the data requires a near-surface electrical conductor and water with some abundance of salts is a good candidate for this internal ocean. Moreover, model calculations show that tidal dissipation in the ice may produce enough heat to keep an ocean underneath an ice lid a few tens of kilometers thick. It is also possible, depending on little known values of 'rheology' parameters and thermal conductivities that radioactive decay and tidal heating in the rock mantle may keep the ocean fluid (Chapter 10.15).

Ganymede has an ice shell about 800 km thick, a 900-km-thick rock mantle shell, and a metallic core with about the same radius. The case for an iron-rich core is particularly appealing for Ganymede because the magnetometers on Galileo have detected a self-generated field on this satellite (see Chapter 10.07). The intrinsic magnetic field of Ganymede is most likely generated in an iron-rich core. The magnetic field data also suggest an induced component although the evidence for the induced field is much less strong. If the latter interpretation is correct, then Ganymede may also have an ocean of perhaps 100 km thickness at a depth where conditions are around the triple point for ice I, III, and water. This will be at a depth of around 150 km for this satellite. Since tidal heating is negligible at present, the heat that keeps the ocean molten must either be derived from radioactive decay in the rock mantle or from heat stored in the deep interior during earlier periods of strong tidal heating (e.g., Malhotra, 1991) or even from accretion. The latter possibilities are highly

speculative, however. In any case, the existence of the self-generated field is evidence enough for a molten core or, at least, a molten core shell.

Callisto is unusual among the Galilean satellites because the gravity data suggest that its interior is not completely differentiated. It is likely that there is an ice shell a few hundreds of kilometers thick overlying a mixed interior of ice and rock also containing iron. It appears as if Callisto has traveled an evolutionary path different from that of Ganymede. The layering may have formed as a consequence of the slow unmixing of the ice and rock/iron components (Nagel *et al.*, 2004). The model requires that Callisto has never been heated above the melting temperature of the ice. It is possible if not likely that the iron is no longer present in its metallic form but is oxidized to form magnetite and fayalite. The oxidation of the iron, if it occurred, would have then precluded the formation of a metallic iron core. As for Europa and Ganymede, the magnetic field data suggest an induced field and an ocean for Callisto as well. As a matter of fact, the evidence is more clear-cut for Callisto than it is for Ganymede. Since Callisto is not in the Laplace resonance, tidal heating can be discarded as a heat source. The remaining candidates are radiogenic heating and accretional heating.

Interior structure models of the largest Saturnian satellite Titan have so far been based on the average density of the body and phase diagrams of water but will improve once two-way Doppler data will be available for this satellite later during the ongoing Cassini mission. The latter mission has provided gravity data for Rhea that suggest that this satellite is not differentiated (Anderson and Schubert, 2007). The shapes of Iapetus and Enceladus (Schubert *et al.*, 2007) have been used to constrain interior models for these satellites (*see* Chapter 10.15 for a discussion). Both satellites are thought to be flattened due to rotation and the flattening increases with concentration of mass in the deep interior. Whether or not the shape of Enceladus is consistent with an differentiated interior is debated although the observed geyser activity and the large luminosity of this satellite indirectly suggest differentiation. The shape of Iapetus is consistent with a higher rotation rate than the present which suggests some orbital evolution that moved the satellite from a previous orbit closer to Saturn to the present one. We note as a caveat that these models assume hydrostatic equilibrium, an assumption that may be questioned given the small size of these bodies.

The terrestrial planets have ice in only modest concentrations, and mostly consist of the rock/iron component. The rock/iron component in these planets, just as in Io, Europa, and Ganymede, is differentiated into a mostly iron core, a silicate rock mantle, and a crust consisting of the low-melting point components of rock (*see* Chapter 10.02). The crust, a layer of some tens to, at most, a few 100 kilometer thickness forms by partial melting of the mantle and melt separation from the mantle. Crustal growth (*see* Chapter 10.09) is an ongoing process on Earth but was most likely almost all completed early on Mars and Mercury.

Seismology has provided us with a detailed image of the interior structure of Earth. It is expected that future missions will provide us with similar data for the other terrestrial planets (*see* Chapter 10.03). The most likely candidate here is Mars and the Moon, for which mission scenarios with seismological networks have been repeatedly studied. The seismological data for Earth show that there are phase transformations mostly at moderate pressures and depths and chemical discontinuities. The existence of these layers, and their thicknesses and depths will vary among the planets. For instance, it is likely that Mercury's mantle will not have major phase-change layers simply because the pressure in this small planet does not reach the levels it reaches within Earth but it may be chemically layered as is the Moon. Although the latter is even smaller than Mercury, limited seismic data available suggest a layered structure. This may be due to the very early evolution (*see* Chapters 10.09 and 10.14) and to the sizes of the planets. For instance, there is evidence that the planets were once covered by what is called magma oceans, rock molten by the dissipation of the energies of infalling planetesimals during the late stages of accretion. The Moon possibly became chemically layered as the magma ocean fractionally crystallized. Earth may have escaped that layering because vigorous convection mixed the products of fractional crystallization. The vigor of convection depends on the size of the body or layer undergoing convection.

10.01.6 Surfaces and Atmospheres

The surfaces of most planets and major-size satellites have been observed optically, in the visible wavelength range of the electromagnetic spectrum, by cameras onboard spacecraft. The gaseous giant planets and subgiants show the top layers of their gaseous envelopes in which clouds may form. Since these

bodies lack solid surfaces, their surface radius has been defined to be the radius at which the atmospheric pressure is 10^5 Pa. Mercury and the major satellites except Titan show their solid surfaces. These bodies are lacking substantial atmospheres. The ability of a planet or satellite to keep an atmosphere depends on the planet's gravity. Equating the kinetic energy of a molecule to its potential energy defines the escape velocity which is independent of the mass of the particle. Atoms, ions, or molecules move at their thermal speed – which obeys a Maxwell–Boltzmann distribution and is proportional to the square-root of the ratio of the temperature to the particle mass – but can be accelerated by other processes much beyond their thermal speeds. Among these processes are impacts, photoionization and pickup by the solar wind, hydrodynamic effects, and sputtering (e.g., Chamberlain and Hunten, 1987). The presence of a magnetic field will help to keep an atmosphere by blocking the solar wind from eroding the atmosphere. Escape due to the thermal speed (at the long end of the Maxwell–Boltzmann distribution) exceeding the escape velocity is called Jeans escape and is most relevant for light species such as hydrogen and helium. Jeans escape can be used for a systematic discussion in terms of planetary mass, radius, and temperature. It explains why Mercury with its comparatively small mass and high surface temperature is prone to losing an atmosphere and why massive Jupiter can bind hydrogen and helium but it cannot explain why Titan has a substantial atmosphere while Ganymede, similar in mass, radius, and temperature, has none. The difference may lie with Titan being rich in ammonia from which the mostly nitrogen atmosphere may have formed by photodissociation (Coutenais and Taylor, 1999 and Chapter 10.15). The compositions of the atmospheres of the terrestrial planets suggest that any solar-type primordial atmosphere has been replaced by a secondary atmosphere that is the result of outgassing and perhaps cometary impacts (e.g., Pepin, 2006). It is also possible that the difference between Titan and Ganymede lies with their early differentiation and outgassing histories.

The surface temperatures of the planets are mostly determined by the solar radiation and therefore decrease with increasing distance from the Sun. The actual surface temperature will depend on the reflectivity or ‘albedo’ in the visible wavelength of the electromagnetic spectrum – on the emissivity in the infrared – and on the thermal and optical properties of the atmosphere. A useful quantity, the ‘effective temperature’, can be calculated assuming thermal

equilibrium of the surface with the solar radiation neglecting the effects of the atmosphere. The effective temperature decreases from 445 K at Mercury to 50 K for Neptune (Table 1). In case of an atmosphere, the energy balance is more complicated. Visible light passes through the atmosphere and reaches the surface where it is in part reflected and in part absorbed and re-radiated. The infrared radiation emanating from the surface is in part absorbed by CO₂ and/or H₂O in the atmosphere and raises the atmosphere temperature. This effect, the so-called greenhouse effect, is particularly important for Venus due to the large amount of CO₂ in its atmosphere. (Other minor constituents contribute to the greenhouse effect.) Here, the surface temperature is about 740 K, about twice the value of the effective temperature. It is possible that Mars had a much denser atmosphere in its early evolution that allowed for higher temperatures through the greenhouse effect and liquid water. This would have helped to develop life on this planet (Chapter 10.12).

The solid surface of Earth is partly covered by clouds in the atmosphere and partly visible depending on the extent and the pattern of the cloud cover. Venus' surface, however, is entirely covered by an optically dense atmosphere and the same is true for the Saturnian satellite Titan. Venus' atmosphere is mostly CO₂ and the atmospheric pressure is 9.2 MPa. The pressure at the bottom of Titan's atmosphere is 150 kPa. The orange color haze that hides the surface from view is most likely smog produced by the photo-dissociation of methane. Venus' surface has been explored by radar observation. Titan is also explored by radar through the Cassini mission (Figure 11) which, at the time of this writing, is orbiting the planet and has repeatedly flown past Titan at a distance of about 1000 km. On 14 January 2005, the European Huygens probe on board Cassini successfully landed on Titan and returned the first images from its surface. The images reveal a complex surface morphology and suggest rivers and perhaps oceans.

The surface of Mars is visible most of the time. The main component of the Martian atmosphere is CO₂ but the atmospheric pressure is much smaller than Venus', of the order of 600 Pa. However, the surface may be hidden during times of global dust storms, that generally occur in late Southern Hemisphere spring. Ancient river beds, outflow channels, and erosion features suggest that the Martian atmosphere was more massive in the past and the climate was wetter and warmer (see Chapter 10.10 for a discussion of atmosphere–surface interactions and Chapter 10.11 for a discussion of water and

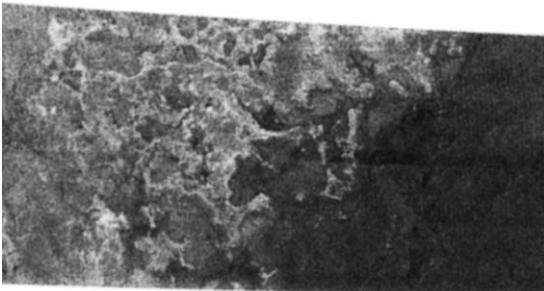


Figure 11 Synthetic aperture radar image of the surface of Titan. The bright, rough region on the left side of the image seems to be topographically high terrain that is cut by channels and bays. The boundary of the bright (rough) region and the dark (smooth) region appears to be a shoreline. The patterns in the dark area indicate that it may once have been flooded, with the liquid having at least partially receded. The image is 175 km high and 330 km wide, and is located at 66° south latitude, 356° west longitude in the southern hemisphere of Titan. ©NASA/JPL.

ice on terrestrial planetary surfaces). The findings of the Mars Exploration rovers clearly show that standing bodies of water on Mars were once present in which minerals like jarosite precipitated. Water is a prerequisite for a planet to be habitable. Habitability of the terrestrial planets is discussed in Chapter 10.12.

The pressure on Mars is close to the triple point pressure of ice, water, and vapor (611 Pa). Thus there is little room for liquid water except for low-lying regions such as the bottom of impact basins like Argyre and Hellas. However, melt water may exist for some time metastably on Mars and longer as its surface freezes over and it is covered by ice (see also articles in Tokano (2005) and references therein). The widely accepted early disappearance of the magnetic field of Mars (see Chapter 10.09 and also see Schubert *et al.* (2000) for a dissenting view) may have been partly responsible for the escape of the early Martian atmosphere (see Dehant *et al.*, 2007 for a recent review). The atmosphere of Earth differs from those of its neighbors. The atmospheric pressure of 0.1 MPa is in between that of the latter two and the main components are N₂, O₂, and H₂O.

The outer surface of Uranus is bland, greenish in color, and mostly featureless. The greenish color is attributed to methane and high-altitude photochemical smog. At the other extreme is Jupiter, whose surface features a large number of bands or stripes largely parallel to the equator interspersed with spots and vortices. Particularly remarkable is the giant red spot, a vortex that covers about 1/10 of the planetary disk (see Figure 1). These features point to a highly dynamic atmosphere. The dynamics is dominated by

the rotation as witnessed by the band structure but the vortices and spots show that these become unstable at various scale lengths. The brownish to reddish color of some of the features is attributed to ammonium-hydrogen-sulfate (NH₄SH) while the whitish colors are attributed to ammonium. Saturn's atmosphere is much calmer than Jupiter's, as the surface patterns suggest. These patterns are stripes similar to those in the Jovian atmosphere but the spots and vortices are missing although wind speeds are extremely high. Neptune is similar to Uranus in that its surface is bluish in color, which is attributed to methane. It resembles Saturn in its stripe pattern. But in addition, there are a few vortices and a large spot that resemble features on Jupiter (the atmosphere dynamics of the giant planets is discussed in Chapter 10.13).

The solid surfaces of the terrestrial planets have some common features. Most prominent among these are craters that occur on a very wide range of sizes (cf. Chapter 10.06). Craters are believed to be remnants of the early evolution of the planets and satellites when the young bodies were bombarded by planetesimals during the late stage of the accretion process. This phase is often termed the phase of heavy bombardment. The large range in size of the craters shows that at that time the size range of bodies in the solar system was likely similar to that of today. The major difference is that the number of bodies was much larger. The larger bodies swept up most of the smaller ones and the impact rate declined with time to its present value, which is small but not zero. The distribution and density of craters on planetary surfaces is an important indication of the age of the surface or parts thereof. A surface is the older the higher the density of impact craters. For instance, Mars shows a dichotomy in surface age (and topography) between its northern lowland and southern highland hemispheres (see Figure 8). The age difference on average is about 1 billion years. Some of the most cratered and therefore oldest surfaces in the solar system are found on Mercury, Earth's Moon, and the Jovian satellite Callisto. There are surfaces that are saturated with craters, implying that for every new crater formed by an impact one existing crater will be destroyed. Geologists and planetary scientists use the term 'exogenic dynamics' to characterize the processes modifying the surface by impacts. Many impact craters on terrestrial planets have been partly or completely destroyed by erosion. Some have been buried by extensive resurfacing in the early history of the planet. On icy satellites long-term relaxation of the ice may cause craters to flatten out.

Planetary surfaces are also modified by erosion (Chapter 10.10) and by processes related to ‘endogenic dynamics’, that is, by processes that originate from within the planet (*see Chapter 10.09*). The most prominent of the latter processes is volcanism. Volcanism is the consequence of (partial) melting of the planetary interior and the rising of the buoyant melt to the surface. It is not restricted to rock and silicate volcanism as we know it from Earth but there may also be ice volcanism on the surfaces of the icy satellites (*see Chapter 10.15*). Volcanism has shaped at least parts of the surfaces of the terrestrial planets and silicate volcanism is the one important element of crustal growth. Prominent volcanic features are the giant Tharsis dome on Mars, the island arcs on Earth, and the maria on the Moon. The surface of Venus is dotted with volcanoes. As we will discuss further below, endogenic activity scales with the mass of a planet. This partly explains why the activity on smaller planets usually dies off earlier than on the larger planets.

A special case is the Jovian satellite Io (see, e.g., [Spencer and Lopes \(2006\)](#)). This satellite is similar in mass to the Moon, yet it is the most volcanically active planetary body in the solar system. Volcanic features cover its surface but impact craters have not been detected. This shows that the surface is very young and is permanently renewed by volcanic material (**Figure 12**). The surface is to a large extent

covered by allotropes of sulfur and by sulfur dioxide. These deposits cause the yellow to whitish color of the surface. The dark spots are likely volcanic vents and sulfur lakes. Sulfur deposits are often found associated with volcanic structures on Earth. The high surface temperatures of up to 1600 K indeed suggest silicate volcanism ([Spencer et al., 2000](#)). The reason for the unusual activity on Io lies with an unusual heat source: Io is flexed by tides raised by its massive primary Jupiter. The deformation energy that is dissipated as heat is sufficient to make it more volcanically active than the largest terrestrial planet, Earth. A further measure of the enormous energy that is dissipated in Io’s interior is the surface heat flow of $2\text{--}3 \text{ W m}^{-2}$, 20 to 30 times larger than the surface heat flow of the Earth. Io’s heat flow is rivaled only by Jupiter which radiates about 5.4 W m^{-2} . The heat flow through the southern polar cap of Enceladus is also exceptionally high, however.

Earth features a style of endogenic activity, plate tectonics, that according to our present knowledge, is unique to Earth (*see Volumes 6 and 7*). Plate tectonics involves the continuous production of basaltic crust along volcanically active linear ridges at the bottom of the oceans. Prominent examples are the Mid-Atlantic Ridge and the East Pacific Rise. It also involves subduction of this crust underneath island arcs and continental margins. Prominent examples are the islands of Japan for the former and the western continental margin of South America for the latter. These subduction zones are the loci of most of the seismic activity of the planet. Both processes cause Earth’s surface to be divided into seven major plates that drift across the surface. Plate tectonics is driven by extremely slowly circulating convection currents in its deep interior. Although convection is not unique to Earth but is expected to occur in the silicate rock shells of the other terrestrial planets as well, this feature is usually hidden underneath a thick stagnant lid (*see Section 10.01.7*). The reason why convection in Earth extends to the surface and includes the crust at the bottom of the oceans is not entirely known. It is speculated that this is caused by the presence of water and there may even be links between plate tectonics and life. In any case, plate tectonics causes the surface of the ocean basins to be completely renewed on a timescale of a few hundred million years. Plate tectonics together with erosion even incorporates recycling of the more stable continents. Material denuded by erosion from the continents is transported to the ocean basins where it is incorporated into the plate tectonics cycle. That

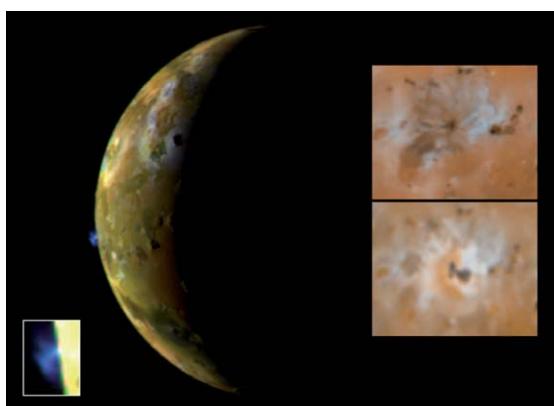


Figure 12 Volcanic plume on Io imaged by the Galileo spacecraft. The blue-colored plume extends to about 100 km above the surface. The blue color is consistent with the presence of sulfur dioxide gas and sulfur dioxide snow condensing as the volcanic gas in the plume expands and cools. The images on the right show a comparison of changes seen near the volcano Ra Patera since Voyager (1979) (top) and Galileo (2001) (bottom). An area of $40\,000 \text{ km}^2$ was newly covered with volcanic material between the two observations. © NASA/JPL.

loss of continental material is balanced by the production of continental rock through volcanic activity. It has been speculated that Mars and Venus went through phases of plate tectonics very early in their histories (*see* Chapter 10.09 and references therein). However, while this is a possibility it remains speculation. Although plate tectonics seems to be unique to Earth, there are other processes of crust recycling. These involve subcrustal erosion as is assumed for Io and the foundering of the surface as has been speculated to have happened on Venus at the sites of some coronae (e.g., Hogenboom and Houseman, 2005 and references therein).

The best-studied example of a single-plate planet is Mars. Two instruments on board spacecraft have recently allowed major progress for detailed studies of the tectonics of this planet. The first is the laser altimeter MOLA on Mars Global Surveyor and the second is the HRSC camera on Mars Express. While MOLA has stopped operating as an altimeter in 2001 due to the finite lifetime of its laser pump diodes, HRSC has covered almost the entire planet at a resolution of 10 m per pixel at the time of this writing. One major achievement by these instruments is the establishment of an accurate geodetic network of surface features.

The topography of Mars exhibits a clear dichotomy which divides the surface into a southern highland hemisphere as heavily cratered as the lunar highlands and rising several thousands of meters above the zero level and a northern lowlands hemisphere that lies well below the datum. The origin of the dichotomy is ascribed variously either to long-wavelength mantle convection (e.g., Zhong and Zuber, 2001) or to postaccretionary core formation sweeping up most of the crustal material into one large proto-continental mass (Davies and Arvidson, 1981). The highlands cover about 60% of the planet, including almost all the southern hemisphere. The surface has survived the heavy bombardment prior to 3.8 billion years ago with only minor modifications and thus records the early history of the planet.

Venus whose surface has been mapped by the Magellan mission with radar is another well-studied single-plate planet. Its hypsometric curve, however, has a single peak rather than the double peak curves of Mars and the Earth. Venus has no distinct volcanic center such as Mars but a prominent rift system, the Beta, Atla, and Themis Regiones. Volcanic features include a large number of comparatively small shield volcanoes that are apparently randomly distributed, volcanic plains and about 500 coronae – circular

large-scale volcano-tectonic features. It is widely accepted that most coronae form above upwelling plumes (Smrekar and Stofan, 1999). The paucity of impact craters on the surface has been used to suggest that Venus was resurfaced about 500–700 million years ago by a volcanic event of global scale (Schaber *et al.*, 1992; McKinnon *et al.*, 1997). This event was followed – as the model suggests – by almost no volcanic activity. The early conclusions from the cratering record have been challenged by, for example, Hauck *et al.* (1998) and Campbell (1999). According to these authors, the cratering record allows a variety of interpretations in terms of volcanic resurfacing including a global decrease in time in the rate of volcanic activity.

10.01.7 Energy Balance and Evolution

Although the evolution of the planets seems to have followed some common general lines, there are significant differences among individual planets. The evolution of the terrestrial planets and moons is discussed in Chapter 10.09 and of the icy satellites in Chapter 10.15. The evolution of the giant planets is discussed in Chapter 10.13. The early evolution of all planets has been dominated by impacts from the debris left over from planet formation (Chapter 10.06). Moreover, isotopic evidence suggests that the planets differentiated by iron-core formation early, within a few tens of million years (Kleine *et al.*, 2002). Of course, the evidence for Mars rests with the very well-founded assumption that the so-called SNC meteorites are in fact rocks from Mars. It is further believed that the planets started hot, heated by the energy deposited in the interior during accretion and by heat dissipated upon differentiation, the terrestrial planets heated to near-melting temperatures of their rock components and to temperatures well above the iron melting temperatures. The planets then cooled from this initial hot state and the cooling drove their evolution. It is possible, for instance, to explain the present rate of infrared radiation from Jupiter simply by cooling. This is not possible for Saturn, for which an additional heat source is required. The continuing gravitational settling of helium may provide enough energy to explain the observed present luminosity of this planet (*see* Chapter 10.13). The energy balances of most other planets and satellites seem to invoke both cooling and heat generated by the decay of radioactive elements, in particular of uranium, thorium, and

potassium (the isotope ^{40}K) which generates 4–5 pW per kg of rock. A further possibility is tidal heating but this mechanism seems to be relevant at present only for Io and, perhaps, Europa and Enceladus.

It should be stressed that the energy balances of the planets are not very well known. A crucial quantity, the surface heat flow – or the intrinsic luminosity – is known for the giant planets, which radiate enough energy in comparison with the solar insolation incident on their surface that their luminosity can be measured from orbit. This is also true for the heat flow from the volcanic hot spots on Io but there is a major uncertainty concerning the flow through the remainder of the surface. Nevertheless, the enormous luminosity of this satellite of at least 1.5 W m^{-2} is remarkable. Another remarkable value is the roughly 6 GW (Spencer *et al.*, 2006) radiated from the southern hemisphere of Enceladus. On the terrestrial planets and the other satellites, as far as we know them today, the surface heat flow must be measured *in situ* (cf. Chapter 7.06). This measurement requires a bore-hole deep enough to avoid the influence of the daily and seasonal temperature variations of the atmosphere. This difficult measurement has been done at many locations on Earth and at two sites on the Moon. In **Table 1** we list the specific luminosities or surface heat flows for the planets and satellites for which the heat flows and luminosities have been measured. Since heat production depends on mass we have divided the heat flows by the masses. A comparison of the entries in **Table 1** shows that Earth and the Moon are within a factor of 2 of the specific radioactive heat-production rate of rock suggesting that about half of the heat flow can be attributed to cooling and half to heat production. This conclusion is known to hold for Earth even on the basis of more specific data and more careful energy balances (cf. Chapters 9.08 and 9.09). An educated prediction of the heat flows for Mars, the other terrestrial planets, and the major satellites could be based on the above observation. The specific luminosity of Uranus also is close to the radiogenic heat production rate per unit mass of rock given above. It is likely that Uranus has a substantial rock core in which that heat is generated.

Neptune, Io, and Europa stand out in comparison with the terrestrial planets. In the case of Neptune, the extra heat flow has been attributed to cooling and whole-planetary contraction. For Io, the extraordinary large heat flow is with little doubt due to tidal heating although ohmic dissipation of energy carried by electric currents between Io and Jupiter is sometimes quoted. The value for Europa in **Table 1** is highly

uncertain and has been derived from indirect observation of tectonic surface features (e.g., Ruiz, 2005). If the estimate were correct then tidal heating would be the best explanation for its value. However, equilibrium models of tidal heating arrive at values smaller by about one order of magnitude.

The dynamic or geologic, chemical, and magnetic evolutions of the planets are mainly governed by their thermal evolution. Planets can be regarded as heat engines that convert heat into mechanical work and magnetic field energy. Mechanical work will be performed by the building of volcanic structures and mountain belts, for example, or during the movement of plates on Earth's surface (cf. Volumes 6 and 7). The agent for these processes is convective heat transport, a mode of heat transport that involves movement of hot material to cold surfaces and cold material to hot surfaces (see, e.g., Schubert *et al.*, 2001 and Chapter 7.02). Convection is driven by sufficiently great temperature differences across a layer or planetary shell and occurs whenever the convective heat-transfer rate exceeds the heat-transfer rate by conduction. Convection is possible even in the solid rock mantles of the planets because of the long timescales that are involved. On geological timescales of tens to hundreds of millions of years, rock behaves as a very viscous fluid. The convection flows at speeds of centimeters per year still transfer more heat than conduction. This is in part due to the enormous masses involved in the flow and in part due to the very low thermal conductivity of rock. There is even a thermostat principle at work in the silicate and ice shells (the Tozer principle; Tozer (1967)) since the viscosities of these solids are strongly temperature dependent. If the flow is too slow to transport the heat the temperature will increase until the flow is strong enough to transfer the heat. If the temperature is so high such that the convective heat transfer rate largely outbalances the heat production rate or the heat flow into the layer from below then the temperature will decrease, viscosity will increase, and convective vigor will decrease. Convection is extensively discussed in Volume 7. The thermal evolution of the Earth is discussed in Volume 9. Chemical differentiation is mostly a consequence of melting and the density difference between melt and solid that will again lead to movement. A particularly good example for the chemical differentiation of the terrestrial planets that occurs in addition to core formation is the growth of the crust through volcanic activity (see Chapter 10.09). However, secondary crustal growth likely involves much longer timescales than core formation.

The material in the crust is derived from partial melting of the mantle. Partial melting is possible when a material of complex chemistry lacks a simple melting temperature. Rather, in materials such as rock (and ice of complex chemistry) there will be a temperature termed the solidus temperature at which the material begins to melt. The melt will consist of that component of the entire assemblage that has the lowest individual melting point. The remaining solid will be depleted in that component. As the temperature rises, other components begin to melt until, finally, at the liquidus 'temperature', the whole assemblage will be molten. In the rock component of planets the component to most easily melt is basalt. Once a basalt liquid forms in the interior of a planet, it will tend to rise to the surface by virtue of its lower density than the density of the remaining rock. In a water–ammonium ice the low melting component will be a water–ammonium mixture of a particular composition. In most cases the melt is likely to be produced within the top few hundred kilometers of a planet by a process termed pressure-release melting. Pressure-release melting occurs when the melting point gradient is steeper than the average temperature gradient. Relatively hot uprising convection currents will emanate from depths where the temperature is below the melting temperature. Upon rising to the surface adiabatically (that is with little heat exchange), the flow will hit a depth where the solidus temperature is exceeded and partial melting will ensue. Because melt is usually much more compressible than solid, there will be a pressure and a depth below which melt will no longer be buoyant to rise to the surface but may actually sink towards the deeper interior. Melting in the interior of planets require heat sources. These are mainly the heat that is stored in the planets during accretion and the heat that is generated by the decay of radiogenic elements. Both are finite reservoirs that will not be replenished. That is why endogenic activity decreases with time.

Since melting occurs when hot material from the deep interior rises in convection currents, the surface distribution of volcanism is speculated to give an indication of the planform of the convection underneath. The timing of the volcanic activity can be used as a guide for assessing the time evolution of the convection. For instance, the differing surfaces of Mars and Venus may have recorded differing planforms of mantle convection and flow history. The dominance of Tharsis on Mars (cf. [Figure 8](#)) suggests that there is or that there once was a giant superplume, a very large upwelling underneath this volcanic dome ([Figure 13](#)). The Mars Global

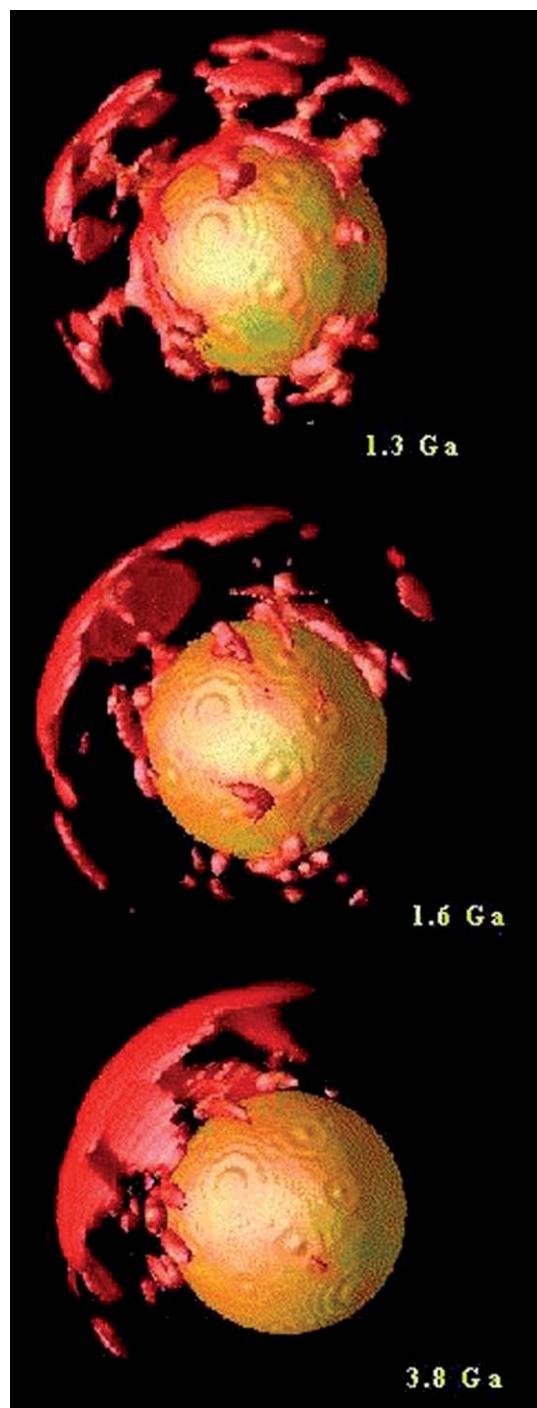


Figure 13 Development of a superplume. This rendering shows super-hot material in the model Martian mantle that rises towards the surface. In this model by Breuer *et al.* (1998) the upwellings concentrate as a function of time into one superplume because of the interaction of the flow with mantle phase transitions. It is likely that superplumes can also result in models with depth-dependent material properties. Courtesy of D. Breuer.

Surveyor mission data suggest that Tharsis formed early and the HRSC data from Mars Express suggest volcanic activity only a few million to ten million years ago. This long-term stability of Tharsis as the major center of volcanic activity on Mars is puzzling, the more so since model calculations suggest that the superplume's lifetime should be not much more than 1 billion years. Although Tharsis apparently formed early, the photogeologic evidence also suggests that volcanic activity on Mars started globally and retreated to Tharsis over time and may then have decreased in vigor. On Venus, there is indication that the volcanic activity has been global even recently (on geological timescales recently involves the past 10–100 million years). There is further (albeit debated) evidence for a global volcanic resurfacing event a few hundred million years ago. This suggests that the thermal history of a planet may have involved episodes of more and of less activity that most likely were linked to episodes of greater and less vigor of convection. Volcanism may also have influences on a hydrosphere and atmosphere by degassing the deep interior of volatile elements. At the time of strong volcanism, an increase of atmospheric gases or water on the planet's surface is expected.

There is a fundamental difference between the evolution of Earth and most other terrestrial planets and major satellites that is related to the occurrence of plate tectonics on Earth. Since the viscosity, or more generally, the rheology of rock is strongly temperature dependent and since the surface temperatures of these planets are much lower than the melting temperatures, it follows that there must be outer layers that are comparatively stiff. These layers are termed the lithospheres. Usually the lithospheres are connected lids that are pierced here and there by volcanic vents. Convection currents flow underneath the lithosphere and deliver heat to the base of the lid through which it is then transferred to the surface by heat conduction. Since these lids are stagnant, this form of tectonics is often termed stagnant-lid tectonics (Figure 14). On Earth, however, the lithosphere is broken into seven major plates that move relative to each other driven by the convective flow underneath. Volcanic activity occurs along some plate margins and results in the growth of plates. Other plate margins are destructive and are loci where plates are forced under their own weight to subduct into the mantle. There is no convincing evidence that plate tectonics occurs on any other planet or satellite although it has been speculated that it may have occurred on early Mars and on the

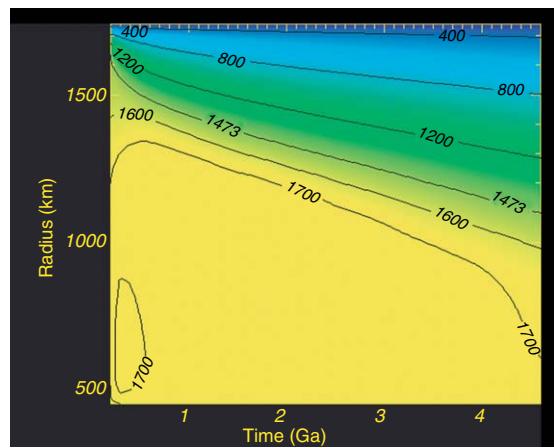


Figure 14 Growth of a stagnant lid in the lunar mantle. The growth of a cold, thermally conductive lid on top of the lunar mantle is shown as resulted from convection model calculations by Konrad and Spohn (1997).

Jovian satellites Europa and Ganymede. There is evidence in the images of the surface units of lateral movement on these two satellites. The Jovian satellite Io does not seem to be undergoing plate tectonics. However, its present resurfacing rate, which is larger than the resurfacing rate of Earth calls for some recycling of Io's crust volcanic material with the underlying mantle. This recycling probably occurs through delamination of the base of the crust. Delamination of the base of the crust may have been or may be operative on other planets such as Venus.

The interior evolution of the gaseous planets is even less well constrained. It must be assumed that the dynamics of their atmospheres is related to the vigor of convection underneath, but the solar insolation also matters. Thus models of the evolution are mostly constrained by their luminosities but as we have seen above these leave enough puzzles to be solved as witnessed by the differences between the luminosities of Neptune and Uranus.

10.01.8 Magnetic Fields and Field Generation

Mercury, Earth, and the giant planets, have largely dipolar magnetic fields that are produced by dynamo action in their interiors. Of the major satellites, only Ganymede is known to produce a magnetic field. The Galileo and Cassini data suggest that neither the other Galilean satellites of Jupiter – in particular

not dynamic Io – nor the big Saturnian Satellite Titan have self-generated fields. The magnetic fields of these planets and moons are described and discussed in Chapter 10.07. The physics of planetary dynamos is discussed in Chapter 10.08.

The icy Jovian satellites Europa and Callisto are surrounded by magnetic fields that vary along with their movement through the magnetosphere of Jupiter (e.g., Kivelson *et al.*, 2004). These fields are, therefore, interpreted to be induced in electrically conducting layers in the satellites' interiors as the satellites orbit Jupiter and are subject to a time-varying magnetic flux. The geometry and the strengths of these fields suggest that the conducting layers are at depths of some tens to a few hundred kilometers and are most likely salty oceans. No evidence for an induced field at Titan has yet been reported. Thermal considerations suggest that these oceans are feasible (see Chapter 10.15). Ganymede may have an induced field on top of its permanent magnetic field although the Galileo magnetometer data are not entirely conclusive in this case. An alternative interpretation associates the signal with higher-order terms in the dynamo field.

Self-sustained magnetic fields are generally thought to be enigmatic to planets, part of their interior evolution during which thermal (and potential energy) is converted into mechanical work and magnetic field energy. The Mars Global Surveyor data have recently confirmed this hypothesis by showing that this planet has a remanently magnetized

crust (Figure 15). This crust must have been magnetized during an epoch when Mars was generating its own magnetic field. (For a recent discussion of Martian magnetism see Connerney *et al.*, 2004 in addition to Chapter 10.07). Interestingly, it is mostly the oldest crust units that are magnetized. It has long been speculated that Mars, Venus, and Earth's Moon once produced magnetic fields by dynamo action in their cores (e.g., Stevenson *et al.*, 1983). The observation of magnetized rock on the surface of Mars now confirms this general notion.

The dynamo mechanism (e.g., Chapter 10.08 but *see also* Chapters 10.09 and 10.13 and chapters in Volume 8, in particular Chapter 8.03) invoked to explain magnetic field generation is similar among the planets although the source regions differ between the terrestrial planets and major satellites, the giant planets, and the subgiants or ice giants, Uranus and Neptune. Required is an electrically conducting and fluid region that undergoes turbulent flow driven by thermal or compositional buoyancy. If a magnetic field exists, for instance the external magnetic field of the Sun or a primary planet, a field will be induced in the source region. The flow will distort the field lines and thereby generate magnetic field energy by induction. If the flow satisfies certain conditions, the interference between the generated field and the preexisting field can be constructive powering and maintaining a magnetic field against dissipative losses. Dynamo theory elucidates the importance of rotation for the twisting of the field

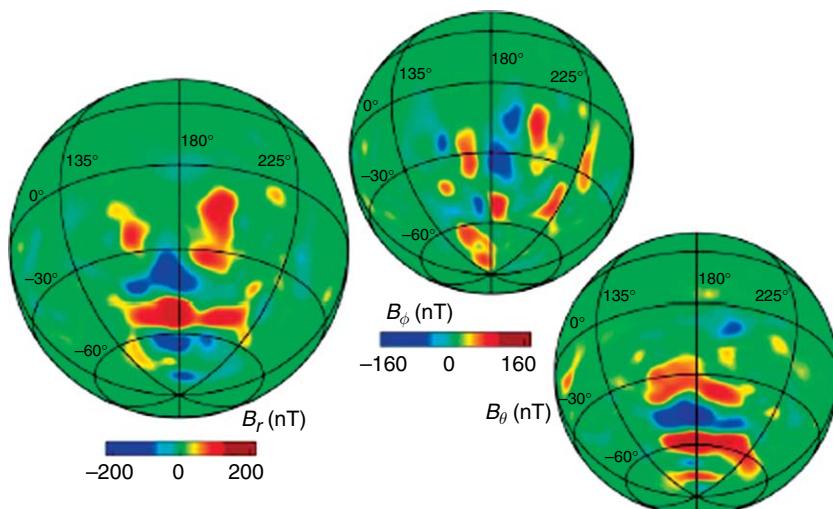


Figure 15 Orthographic projections of the three components of the magnetic field in spherical coordinates (r , radius; Θ , latitude; and Φ , longitude) onto the surface of Mars at a nominal mapping altitude of 400 km. From Connerney JEP, Acuna MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Review* 111: 1–32.

lines and the induction. Indeed, balancing the major dissipation mechanism, Ohmic dissipation, or Joule heating, against rotational inertial forces gives a scaling law for planetary magnetic fields that is supported by the observations. In recent years, numerical dynamo models have been published in increasing numbers. A most recent model of the Hermean dynamo is by [Christensen \(2006\)](#).

For the terrestrial planets and the major satellites the candidate source regions are the fluid metallic cores or outer-core shells. Because of their low viscosities (around 1 Pa) these regions are often unstable with respect to thermal convection. The stability is believed to depend on the rate of heat removal from the core by the mantle convection flow. If the latter is too low (below about 10 mW m^{-2}), the heat flow in the mantle can be balanced by heat conduction in the core and the core will be stably stratified. Dynamo action is then not possible. If the heat flow into the mantle is larger than the critical value, convection in the core must be invoked to balance the heat removal rate. The situation becomes more favorable for convection in the core if the core begins to freeze and grow a solid inner core. First, latent heat liberated upon inner-core growth helps to power the dynamo. Second, the core may contain sulfur and/or oxygen as light alloying elements. The light alloying elements expelled from the solid inner core can drive (chemical) convection and a dynamo very effectively because this dynamo will not be subject to a Carnot efficiency factor that by the second law of thermodynamics limits the efficiency at which a heat engine can do work. It is believed that the cooling of the core and freezing of the inner core drive Earth's core dynamo.

Dynamo action in a terrestrial planet thus depends on the efficiency of mantle convection, on the composition of the core, and the core material phase diagram. Plate tectonics, a mode of mantle convection associated with present-day Earth is very effective at cooling the deep interior of a planet. Because the cold plates sink through the mantle, the deep mantle and core will cool effectively. It is therefore conceivable that Earth's core has been cooling to temperatures below the liquidus of its core alloy. Venus, a planet of similar size, appears to have been lacking that efficient cooling mechanism and appears to be cooling by convection underneath a stagnant lid. Consequently, the heat flow from the core is subcritical and during the evolution of the planet the core has not been cooling enough to reach liquidus temperatures. As a consequence of this and the general decline of mantle convective vigor, the core

became stably stratified and a possible early dynamo ceased to operate. A similar scenario is likely for Mars. Early plate tectonics may have helped in generating the early field that has long become extinct but is recorded in the remanently magnetized regions of the crust (Chapter 10.09). There is little hope to find remanently magnetized crust on Venus should landing missions be able to overcome the forbidding operating conditions on this planet. The surface temperature of around 450°C is above the temperatures at which candidate minerals become remanently magnetized. The Moon may be telling basically the same story although here the evidence for early magnetic field generation may have partly been lost. There is doubt though that the small lunar core could have supported a dynamo. Alternative explanations for the recorded remanently magnetized crust units invoke plasma clouds generated by major impacts. Some of the strongest magnetic anomalies (albeit weak in comparison to Mars) are suspiciously located at the antipodes of major impact basins. Mercury differs from both the Moon and Mars because of its extraordinarily large core and thin rocky mantle. The thin mantle may be quite effective, as model calculations suggest ([Schubert *et al.*, 1988](#)), at removing core heat and driving a dynamo that produces a magnetic field, albeit weak. The weakness of the field has been discussed to be difficult to explain with conventional dynamo theory. An alternative dynamo mechanism invokes the thermo-electric effect ([Stevenson, 1987](#)). The recent numerical model by [Christensen \(2006\)](#) invokes the skin effect in the outer layers of the Hermean core to reduce the strength of the field observed at or above the surface. Ganymede is puzzling because it still produces a magnetic field although a comparison with Mars and the other terrestrial planets would have suggested that its dynamo should also have ceased early on. Perhaps Ganymede formed its core relatively slowly and late and we may be looking at a core in a phase of its evolution that those of Mars and Moon passed through during the first billion years of their histories. But much of the heat of core formation would be dissipated if the core formed over billions years instead of a few tens of million years as in Mars or the Earth. Thus, if Ganymede formed its core slowly it is not clear that it would be sufficiently superheated with respect to the mantle to drive thermal convection. Ganymede may differ from Mars and Venus by having a wetter mantle with a softer rheology in which case its core would be cooled more effectively by more vigorous mantle convection. In

analogy with the Earth, and accepting that water is instrumental for plate tectonics on Earth (see Volume 6) it can be speculated that some form of plate tectonics works in Ganymede's rock mantle underneath the ice cover. The surface morphology of Ganymede at least suggests some form of drift of ice blocks there.

Dynamo action in the gaseous giant planets (see Chapters 10.08, 10.09, and 10.13) most likely happens in the regions where hydrogen becomes metallic due to the extraordinary large pressure. The magnetic fields of these planets are most likely powered by the energy that has been stored in the planets during accretion and that is removed by convection. The transition to metallic hydrogen occurs at a depth of roughly 15 000 km in Jupiter and almost twice as deep in Saturn. The larger depth of dynamo action may partly explain why Saturn's field is much more ideal in terms of a dipole than Jupiter's. In the source regions, the fields are likely to be very complex and not at all similar to a dipolar planetary field. The field in the source region can be thought of as a superposition of many multipolar fields of various amplitudes. Distance from the source region affects the higher order multipolar fields much more strongly than the lower order fields. If n is the order ($n=1$ for a dipole) then the strength of the field components decreases with radial distance r from the source regions as $r^{-(n+1)}$. Thus, a planet with a comparatively deep source region should feature a more ideal dipolar field as compared with one where the source region lies less deep. This does not explain entirely the characteristics of the Saturnian magnetic field for which screening by helium rain has been additionally invoked.

Pressure in the interiors of the subgiants Uranus and Neptune is not sufficient to cause a transition to metallic hydrogen. Rather, it is believed that there are ionic oceans at relatively shallow depths in these planets. In these ionic oceans the magnetic fields can be generated. The fields of the subgiants have been measured only by a single flyby each and are therefore not well known. Nevertheless, it appears that these fields are very complex in their topologies. This fits in nicely with the idea of the fields being generated at relatively shallow depths.

10.01.9 Origin of the Solar System

The origin of the solar system is still not completely understood. Formation of planets has recently been reviewed by, for example, [Boss \(2002\)](#), for giant and subgiant planets, by [Bodenheimer and Lin \(2002\)](#)

including evidence from extrasolar planets, and by [Chambers \(2004\)](#) for terrestrial planets. See also the compilations *Protostars and Planets IV* ([Mannings et al., 2000](#)) and *V* ([Reipurth et al., 2007](#)). The formation of satellites is discussed in Chapter 10.14. The basic elements of the theory are similar to the Kant–Laplace hypothesis of formation from a gaseous nebula but the details are debated. It is widely accepted that the nebula collapsed to form a central mass concentration – the proto-sun – surrounded by a spreading thin gaseous disk. The most widely held view postulates that temperature in the inner part of the nebula soon became low enough – about 1500 K – to allow the condensation of silicate and iron grains. In the outer solar system, beyond about 5 AU, temperature became so small as to allow condensation of ice phases. Within the first few million years, these solid grains aggregated to form bigger grains that, in turn, aggregated to form even bigger ones. The cascading scenario of ever bigger and fewer planetesimals led to the formation of solid bodies the size and mass of the terrestrial planets in the inner solar system and a few tens of Earth masses in the outer solar system. In the outer solar system these solid proto-planets became the cores of the future giant planets. Growth of the solid proto-planets slowed to almost a standstill as the solid matter became exhausted, at least in the feeding zone of the proto-planets which is determined by the competition between a proto-planet's gravity and the gravitational pull of the Sun. At about one Earth-mass a proto-planet begins to accrete gas onto its surface. This happened at a time when in the outer solar system the solid cores were still accreting solid planetesimals. The heat dissipated during the influx of planetesimals was radiated from the surface of the forming gaseous envelop. After the influx of solid matter ceased, the proto-giant began to cool and to contract thereby increasing its potential to accrete more gas. The growing mass led to even more accretion and runaway growth was set in place. Runaway accretion came to a halt when a gap formed in the nebula around the planet. This gap can form as a consequence of the competing gravities of the proto-planet and the Sun and as a consequence of the finite viscosity of the gas.

This model offers an explanation of the grand features of the solar system: The inner planets are solid and of refractory composition because the temperature in the inner nebula favored the condensation of refractory phases. The inner planets are small because the feeding zone was smaller with less mass of solid particles. The bigger of the inner planets – Earth and Venus – have atmospheres with masses as expected

although the present atmospheres are not likely to be the primordial ones which were lost and replaced by degassing of the interiors. The present atmospheres are therefore gas that was originally deposited as solids in the growing proto-planet. In the outer solar system we find planets with cores of some 10 Earth masses and massive gaseous envelopes. In addition, we find wealths of satellite systems witnessing accretion in orbit around the growing proto-giants. There are a number of features that can be explained by fine-tuning the theory, for instance, the differences between Jupiter and Saturn and Uranus and Neptune but this will not change the grand picture. A further consideration is the formation of the asteroid belt and the formation of relatively small Mars which must be attributed to nearby massive Jupiter. Giant impacts, a mechanism of great importance for some bodies are the final events of the accretion scenario. Giant impacts are thought to have caused the formation of the Moon (e.g., Cameron, 1997; Chapter 10.14) and may be responsible for the high density of Mercury. A giant impact is a collision of almost grown proto-planets. For instance, it is thought that the Moon formed after a Mars-sized proto-planet hit the proto-Earth. The outer layers of the Earth vaporized during the impact and the Moon formed from the condensed vapor cloud. Among the arguments for the giant impact hypothesis for the formation of the Moon is the geochemical closeness of the bulk Moon to the Earth's mantle. Mercury may have suffered a similar giant impact that removed a substantial part of the original planetary mantle. This may explain why present Mercury has a comparatively big core and a thin mantle. A Moon did not form around Mercury because of the closeness to the Sun. Giant impacts may also affect the rotation and it has been proposed that Venus' retrograde rotation may thus be explained. Similar explanations have been brought forward for Uranus.

10.01.10 Concluding Remarks

With this overview we have tried to take the reader on a tour through the solar system and provide the stage for the detailed articles in the remainder of the volume. The planets are largely of solar composition but differ in their depletion in volatile elements. The degree of depletion increases with decreasing mass and with decreasing distance from the Sun with Jupiter being closest in composition to the Sun and with the terrestrial planets being mostly depleted in volatile elements. The planets are mostly internally

differentiated with the heavy elements tending to be found near the center and the most volatile elements near the surface. The gaseous planets present the observer with their top layers of their envelopes while most of the smaller planets present their solid surfaces. Relicts of very early impacts characterize these surfaces to varying degrees. But these surfaces also reveal the signs of volcanic and tectonic activity, and erosion. In general, the smaller the solid-surface planet, the more characterized is its surface by early impacts. The volcanic activity and the magnetic fields of those planets that have self-generated fields are due to convection currents in their interiors driven by cooling and by heat generated by radioactive decay or by the dissipation of tidal energy. The source regions of the magnetic fields are metallic or, in the cases of Uranus and Neptune, ionic fluids. Life has had a chance to develop on Earth but possibly also on Mars and Europa (Chapter 10.12). Finding evidence for life outside Earth is a challenge for the future. The bio-geo system of Earth may be a self-regulating system that sets this planet apart from its sister planets. Its position in the solar system, together with a moderate greenhouse effect in its atmosphere, has provided the planet with habitable temperatures and with an atmospheric pressure allowing for liquid water. Life plays an important role here by removing CO₂ from the atmosphere and moderating the greenhouse effect. Water, on the other hand, may be instrumental for plate tectonics since its effect on the rheology may pave the way for surface plates to subduct to the deep interior, thereby cooling the core. The cooling of the core is speculated to be instrumental for magnetic field generation. The magnetic field is a feature that Earth's brother and sister planets Mars and Venus lack. The magnetic field protects the environment, and thereby life, from radiation. It will be important to see whether or not life developed in the less friendly environments of Mars or even Europa or early Venus.

The future in scientific planetary exploration on the one hand certainly lies with systematic remote sensing with long-term orbiters. These missions will create data volumes that either require long-term operation and communication or significantly increased data acquisition and transmission rates. On the other hand, *in situ* exploration with rovers and planes and networks of surface stations with long operational life times to observe the atmospheres and the interiors will be important. A further challenge for the future will be the return of samples from planets and small bodies. The sophistication of terrestrial laboratories cannot simply be transferred to other planetary surfaces

given the scarcity of resources such as mass and power on spacecraft. The technological challenge remains miniaturization. Although many planetary worlds are forbidding to human exploration and require the operation of robots (think of Venus), it is probably true that human explorers cannot be topped in their ability for *in situ* scientific study and operation of hardware.

Closer to home the challenges for planetary geo-physics and planetary science in general lie with numerical modeling and laboratory studies. For instance, the rheology of planetary matter is not sufficiently understood as are the properties of materials in the deep interiors of the Earth and Venus, let alone Jupiter and the other giant planets. Numerical modeling has advanced largely in step with the advancement of computer power but for geophysical modeling of, for example, the dynamo and mantle convection the achievable parameter ranges are still in many cases far from the realistic ones.

Exploration has started to look beyond the solar system for other planetary systems. More than 100 extra-solar planets have been discovered to date that are in the mass range of about 1/10 to 10 Jupiter masses, or, assuming the same density, roughly one half to two times the size of the latter. Of course, interest is directed towards Earth-like planets in other solar systems! These planets are beyond the reach of present observational tools and space agencies are planning to launch space-based telescopes that have more favorable seeing conditions such as the European Corot satellite launched in December of 2006 and the NASA terrestrial planet finder. But the question is, is the Earth typical? May other solar systems have planets with thick ice covers such as the icy moons of Jupiter, but bigger? Is it conceivable that there are planets of about Earth's size with primordial, Jupiter-like atmospheres (Stevenson, 2004) From a theoretical point of view, these worlds are possible (a statement which may simply reflect our ignorance). In the icy planets, radiogenic heat may support oceans and it is conceivable that these oceans as well as the primordial gaseous envelopes may harbor life.

References

- Acuca MH, Connerney JEP, Ness NF, et al. (1999) Global distribution of crustal magnetism discovered by the Mars Global Surveyor MAG/ER experiment. *Science* 284: 790–793.
- Anderson JD and Schubert G (2007) Saturn's satellite Rhea is a homogeneous mixture of rock and ice. *Geophysical Research Letters* 34: L02202 (doi: 10.1029/2006GL028100).
- Bagenal F, Dowling TE, and McKinnon WB (eds.) (2004) *Jupiter*. Cambridge: Cambridge University Press.
- Basri G and Brown ME (2006) Planetesimals to brown dwarfs: What is a planet? *Annual Review of Earth and Planetary Science* 34: 193–216.
- Bodenheimer P and Lin DNC (2002) Implications from extrasolar planets for understanding planet formation. *Annual Review of Astronomy and Astrophysics* 30: 113–148.
- Boss AP (2002) Formation of gas and ice giant planets. *Earth and Planetary Science Letters* 202: 513–523.
- Breuer D, Yuen DA, Spohn T, and Zhang S (1998) Three dimensional models of Martian mantle convection with phase transitions. *Geophysical Research Letters* 25: 229–232.
- Cameron AGW (1997) The origin of the Moon and the single impact hypotheses V. *Icarus* 126: 126–137.
- Campbell BA (1999) Surface formation rates and impact crater densities on Venus. *Journal of Geophysical Research* 104: 21951–21955.
- Chambers JE (2004) Planetary accretion in the inner solar system. *Earth and Planetary Science Letters* 223: 241–252.
- Chamberlain JW and Hunten DM (1987) *Theory of Planetary Atmospheres*, 2nd edn., 480 pp. New York: Academic Press.
- Christensen U (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 444: 1056–1058.
- Coustonis A and Taylor F (1999) *Titan: The Earth-like Moon*, 344 pp. London: World Scientific Publishing.
- Connerney JEP, Acuna MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Review* 111: 1–32.
- Davies GF and Arvidson RE (1981) Martian thermal history, core segregation, and tectonics. *Icarus* 45: 339–346.
- Davis AM (ed.) (2004) *Treatise on Geochemistry*, vol. 1. Amsterdam: Elsevier.
- Dehant V, Lammer H, Kulikov YuN, et al. (2007) The planetary magnetic dynamo effect on atmospheric protection of early Earth and Mars. *Space Science Review* (in press).
- Friedman L and Kraemer R (1999) Planetary exploration missions. In: Weissmann PR, McFadden L, and Johnson TV (eds.) *Encyclopedia of the Solar System*, pp. 923–940. New York: Academic Press.
- Guillot T (2005) Interiors of giant planets: Models and outstanding questions. *Annual Review of Earth and Planetary Science* 33: 493–530.
- Hogenboom T and Houseman GA (2005) Rayleigh Taylor instability as a mechanism for coronae formation on Venus. *Icarus* 180: 292–307.
- Hauck SA, II, Phillips R, and Price MH (1998) Venus: Crater distribution and plains resurfacing models. *Journal of Geophysical Research* 103: 13635–13642.
- Hubbard WB and Marley MS (1989) Optimized Jupiter, Saturn, and Uranus interior models. *Icarus* 78: 102–118.
- Kivelson MG, Bagenal F, Kurth WS, Neubauer FM, Paranicas C, and Sauer J (2004) Magnetospheric interactions with satellites. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter*, pp. 513–536. Cambridge: Cambridge University Press.
- Kleine T, Münker C, Mezger K, and Palme H (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf-W chronometry. *Nature* 418: 952–955.
- Konrad W and Spohn T (1997) Thermal history of the moon: Implications for an early core dynamo and post-accretional magmatism. *Advances in Space Research* 19: 1511–1521.
- Lebreton J-P, Witasse O, Sollazzo C, et al. (2005) An overview of the descent and landing of the Huygens probe on Titan. *Nature* 438: 758–764.
- Lodders K and Fegley B, Jr. (1998) *The Planetary Scientist's Companion*, 371 pp. New York: Oxford University Press.

- Malhotra R (1991) Tidal origin of the Laplace resonance and the resurfacing of Ganymede. *Icarus* 94: 399–412.
- Mannings V, Boss AP, and Russel SS (eds.) (2000) *Protostars and Planets IV*, 1422 pp. Tucson, AZ: Arizona University Press.
- McKinnon WB, Zahnle KJ, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Models and observations. In: Bouger SW, Hunter DM, and Phillips RJ (eds.) *Venus II*. Arizona: The University of Arizona Press.
- Nagel K, Breuer D, and Spohn T (2004) A model for the interior structure, evolution, and differentiation of Callisto. *Icarus* 169: 402–412.
- Neukum G, Jaumann R, Hoffmann H, et al. (2004) Recent and episodic volcanic and glacial activity on Mars revealed by the high resolution stereo camera. *Nature* 432: 971–979.
- Peale SJ (1988) The rotational dynamics of Mercury and the state of its core. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 461–493. Tucson, AZ: University of Arizona Press.
- Pepin RO (2006) Atmospheres on the terrestrial planets: Clues to origin and evolution. *Earth and Planetary Science Letters* 252: 1–14.
- Porco CC, Baker E, Barbara J, et al. (2005) Cassini Imaging Science: Initial results on Phoebe and Iapetus. *Science* 307: 1237–1242.
- Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active south pole of Enceladus. *Science* 311: 1393–1401.
- Ruiz J (2005) The heat flow of Europa. *Icarus* 177: 438–446.
- Reipurth B, Jewitt D, and Keil K (eds.) (2007) *Protostars and Planets V*, 1024 pp. Arizona University Press.
- Schaber GG, Strom RG, Moore HJ, et al. (1992) Geology and distribution of impact craters on Venus: What are they telling us? *Journal of Geophysical Research* 97: 13257–13301.
- Schubert G, Olsen P, and Turcotte DL (2001) *Mantle convection in the Earth and Planets*, 956 pp. Cambridge: Cambridge University Press.
- Schubert G, Ross MN, Stevenson DJ, and Spohn T (1988) Mercury's thermal history and the generation of its magnetic field. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 429–460. Tucson, AZ: University of Arizona Press.
- Schubert G, Russell CT, and Moore WB (2000) Timing of the Martian dynamo. *Nature* 408: 666–667.
- Schubert G, Anderson JD, Spohn T, and McKinnon WB (2004) Interior composition, structure and dynamics of the Galilean satellites. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter*, pp. 35–57. Cambridge: Cambridge University Press.
- Schubert G, Anderson JD, Travis BJ, and Palguta J (2007) Enceladus: Present internal structure and differentiation by early and long-term radiogenic heating. *Icarus* 188: 345–355.
- Shirley JH and Fairbridge RW (eds.) (1997) *Encyclopedia of Planetary Sciences*, 990 pp. London: Chapman and Hall.
- Smith DE, Zuber MT, Neumann GA, et al. (1999) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1503.
- Smrekar SE and Stofan ER (1999) Evidence for delamination of corona-dominated topographic rises on Venus. *Icarus* 139: 100–115.
- Sohl F, Spohn T, Breuer D, and Nagel K (2002) Implications from Galileo observations on the interior structure and chemistry of the Galilean satellites. *Icarus* 157: 104–119.
- Sotin C, Jaumann R, Buratti BJ, et al. (2005) Release of volatiles from a possible cryovolcano from near-infrared imaging of Titan. *Nature* 435: 786–789.
- Spencer JR, Rathbun JA, Travis LD, Tamppari LK, Barnard L, and Martin TZ (2000) Io's thermal emission from the Galileo photopolarimeter-radiometer. *Science* 288: 1198–1201.
- Spencer JR, Pearl JC, Segura M, et al. (2006) Cassini encounters Enceladus: Background and the discovery of a south polar hotspot. *Science* 311: 1401–1405.
- Spencer JR and Lopes R (2006) *Io after Galileo: A New View of Jupiter's Volcanic Moon*, 342 pp. Berlin: Springer.
- Spohn T (1997) Tides of Io. In: Wilhelm H, Zurn W, and Wenzel HG (eds.) *Lecture Notes in Earth Sciences Vol. 66: Tidal Phenomena*, pp. 345–377. Heidelberg: Springer-Verlag.
- Squyres SW, Arvidson RE, Baumgartner ET, et al. (2003) Athena Mars rover science investigation. *Journal of Geophysical Research* 108: 8062 (doi:10.1029/2003JE002121).
- Squyres SW and Knoll AH (eds.) (2005) Sedimentary geology at Meridiani Planum, Mars. *Earth and Planetary Science Letters* 240: 122–148.
- Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism of the terrestrial planets. *Icarus* 54: 466–489.
- Stevenson DJ (1987) Mercury's magnetic field: A thermoelectric dynamo? *Earth and Planetary Science Letters* 82: 114–120.
- Stevenson DJ (2004) Planetary diversity. *Special Issue: Planetary Diversity. Physics Today* 43–48.
- Tokano T (ed.) (2005) *Water on Mars and Life*, 331 pp. Heidelberg: Springer.
- Tozer DC (1967) Towards a theory of thermal convection in the mantle. In: Gaskell TF (ed.) *The Earth's Mantle*, pp. 325–353. London: Academic Press.
- Weissmann PR, McFadden L, and Johnson TV (eds.) (1999) I, 992 pp. New York: Academic Press.
- Yoder CF and Peale SJ (1981) The tides of Io. *Icarus* 47: 1–35.
- Zhong S and Zuber MT (2001) Degree 1 mantle convection and the crustal dichotomy on Mars. *Earth and Planetary Science Letters* 189: 75–84.

Relevant Websites

- <http://ssd.jpl.gov> – Jet Propulsion Laboratory Solar System Dynamics.
- <http://nssdc.gsfc.nasa.gov> – National Space Science Data Center.
- <http://photojournal.jpl.nasa.gov> – Planetary Photojournal.

10.02 Interior Structure, Composition, and Mineralogy of the Terrestrial Planets

F. Sohl, DLR Institute of Planetary Research, Berlin, Germany

G. Schubert, University of California, Los Angeles, CA, USA

© 2007 Elsevier B.V. All rights reserved.

10.02.1	Introduction	28
10.02.2	Observational Methods	29
10.02.2.1	Geodesy	29
10.02.2.2	Rotation and Tides	30
10.02.2.3	Gravity and Topography	31
10.02.2.4	Magnetic Fields	32
10.02.2.5	Electromagnetics	32
10.02.2.6	Seismology	33
10.02.2.7	Surface Geology and Composition	34
10.02.2.8	Material Properties	34
10.02.3	Interior Structure and Composition	35
10.02.3.1	Two- and Three-Layer Structural Models	35
10.02.3.2	Multilayer Structural Models	36
10.02.3.2.1	Governing equations	36
10.02.3.2.2	Equation of state	37
10.02.3.2.3	Boundary conditions	39
10.02.3.2.4	Numerical solution	39
10.02.4	Earth as a Type Example of a Terrestrial Planet	40
10.02.4.1	General	40
10.02.4.2	Interior Structure	40
10.02.4.3	Composition	41
10.02.4.4	Mineralogy	41
10.02.5	The Moon	42
10.02.5.1	General	42
10.02.5.2	Interior Structure	42
10.02.5.3	Composition	43
10.02.5.4	Lunar Crust	43
10.02.5.5	Lunar Seismology and Mineralogy	44
10.02.6	Mercury	47
10.02.6.1	General	47
10.02.6.2	Interior Structure	48
10.02.6.3	Composition	49
10.02.6.4	Magnetic Field	49
10.02.6.5	Future Exploration	50
10.02.7	Mars	51
10.02.7.1	General	51
10.02.7.2	Interior Structure	51
10.02.7.3	Composition	54
10.02.7.4	Mineralogy	55
10.02.7.5	Martian Seismicity	57
10.02.8	Venus	57
10.02.8.1	General	58
10.02.8.2	Interior Structure	58

10.02.8.3	Composition	58
10.02.8.4	Tectonism	59
10.02.8.5	Dynamics	59
10.02.9	Summary and Outlook	60
References		61

10.02.1 Introduction

The inner planets Mercury, Venus, Earth, and Mars and terrestrial-type bodies like the Moon and some of the outer planet satellites are mainly composed of silicate rock and metals like iron. They are characterized by relatively small masses and radii, and large densities in comparison to the giant planets of the outer solar system. Measurements of rotation and gravitational and magnetic fields indicate that the interiors of these bodies, like the interior of the Earth, are chemically and rheologically layered and subdivided into iron-rich cores, silicate mantles, and rocky crusts derived from partial melting of the mantles. It is generally believed that the internal differentiation of the terrestrial planets and major satellites took place early in their histories only shortly after accretion from colliding planetesimals (Kleine *et al.*, 2002). Even the asteroid 4 Vesta, an accretional remnant left over from that period reveals a variegated surface with characteristic spectral features hinting at a strongly differentiated interior (Righter and Drake, 1996, 1997; Ghosh and McSween, Jr., 1998; Drake, 2001; Keil, 2002; Kleine *et al.*, 2002).

The investigation of planetary interiors is among the most important scientific objectives of interplanetary space missions. The internal structure and bulk composition of terrestrial planetary bodies provide important clues on the origin and early evolution of the solar system. Many large-scale planetary processes are controlled by the internal structure of these bodies. Surface geology and tectonic features are mainly affected by mechanisms that dominate the transport of internal heat from the interior to the surface. The existence of self-sustained and/or induced magnetic fields requires reservoirs of electrically conducting fluids at some depth, providing additional constraints on the present thermal state of these bodies. Since a fluid layer within a planetary body mechanically decouples the deep interior from its outer portion, the propagation of seismic waves and the way in which a planet or

satellite responds to tides is strongly affected by the physical state of its interior.

Models of the interior structure of a number of terrestrial-type planetary bodies are shown in [Figure 1](#). These models are based on theoretical considerations which are strongly constrained by Earth-based and remote-sensing observations, *in situ* measurements, and laboratory studies. Interior structure models aim at calculating (1) the volumes and masses of major chemical reservoirs that contribute to the bulk composition; (2) the depths to chemical and rheological discontinuities and mineral phase boundaries; and (3) depth variations of pressure, temperature, density, and composition. In the absence of seismological data, the most important parameter that permits a rough estimate of how the interior is composed is the average density. Mean density is affected by self-compression, thermal expansion, and pressure-induced mineral phase transitions caused by the weight of overlying layers. In lieu of seismological observations, the determination of a planet's axial moment of inertia, usually by the measurement of the precession rate of the spin axis, provides the principal constraint on the concentration of mass toward the center of the planet. Large-scale gravity and topography data are also important in constraining internal mass distributions since the shapes of the physical surface and the external gravitational field are tied to the radial density distribution and compositionally and/or thermally induced lateral density heterogeneities.

This chapter is arranged as follows: in Section 10.02.2, those observational methods that provide constraints on the interior structure and composition of terrestrial planetary bodies are reviewed. Section 10.02.3 provides the mathematical background for the construction of spherically symmetric density models and discusses thermodynamic equation of state (EOS) parameters relating local density to the prevalent ambient pressure and temperature conditions. Section 10.02.4 describes the interior of the Earth as a type example of a terrestrial planet and is followed by Sections 10.02.5–10.02.8 characterizing in some detail

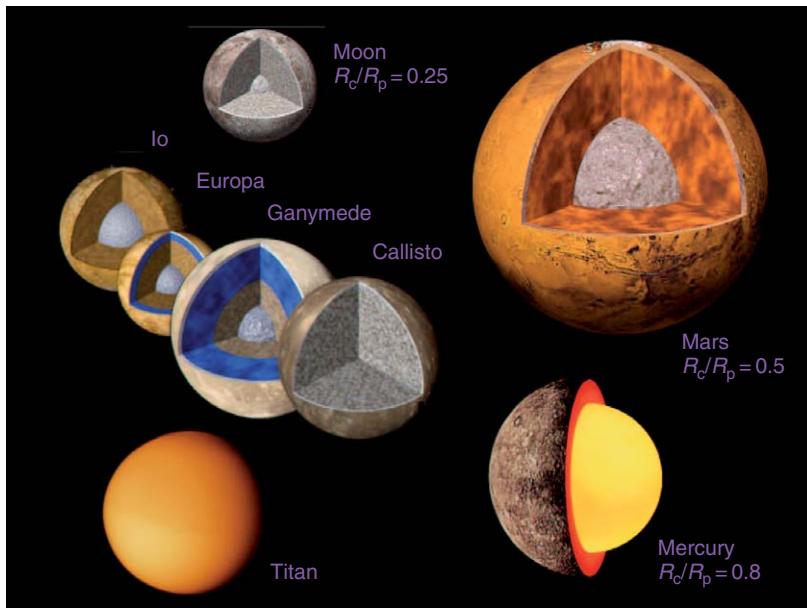


Figure 1 Cut-away views of the interiors of a number of terrestrial-type bodies. R_c/R_p denotes the core radius of a planet or satellite relative to its total radius.

the interiors of the Moon, Mercury, Mars, and Venus. Finally, a summary of this chapter is included in Section 10.02.9.

10.02.2 Observational Methods

Important observational constraints on planetary interiors are provided by astronomical and geodetic methods, rotational and tidal variations, relationships between gravity and topography, magnetic field observations, planetary seismology, interpretations of geologic surface features and compositional variations, and laboratory studies of planetary materials and meteorites.

10.02.2.1 Geodesy

Bulk composition and interior structure of a planet or satellite are mainly constrained by its mean density and mean moment of inertia. The mean density of a planetary body is calculated from its mass and volume. To that end, it is necessary to infer the size and global shape of a planet or satellite from images acquired telescopically using micrometric or photographic techniques or by cameras onboard spacecraft. The global shape or figure of a body depends on its mass, size, rotational and tidal state, and material

strength. Detailed topographic maps of bodies like the Moon, Mars, and Venus have been obtained by using Earth-based radar observations (Muhleman *et al.*, 1995; Margot *et al.*, 1999; Harmon *et al.*, 2001), radio-occultation data, and radar and laser altimetry from orbiting spacecraft (Zuber *et al.*, 1992, 1994; Smith *et al.*, 1999a; Rappaport *et al.*, 1999). Future missions to Mercury also will benefit from laser altimetry (Solomon, 2003). Though the masses of planets and satellites have been known from astronomical observations for a long time, the most current and accurate values are based on measurements of spacecraft that landed, orbited, or flew by.

The gravitational field of a planet or satellite is closely related to the distribution of mass inside the body. Doppler tracking of a spacecraft using its radio communications signal determines its orbit or trajectory from which the gravitational field and the planetocentric constant GM (where G is the universal gravitational constant and M is total mass) can be inferred. The Doppler tracking method is highly accurate in the case of orbiting spacecraft, so that the nominal error of the deduced value of the mean density of the body, for example, $3933.5 \pm 0.4 \text{ kg m}^{-3}$ in the case of Mars, is primarily due to uncertainties in the value of G (Esposito *et al.*, 1992). The mean density is the main indicator of the bulk composition of the terrestrial planets and terrestrial-type bodies.

The density of a material is strongly dependent on porosity and ambient pressure and temperature conditions the latter of which may rise to a few hundreds of GPa and thousands of kelvins deep inside large planets like Venus and Earth. Uncompressed densities vary from about 1000 kg m^{-3} for planetary ices, to 3000 kg m^{-3} for silicate rocks, and up to 8000 kg m^{-3} for metallic iron. The densities of the Moon, the Jovian satellite Io, and the inner terrestrial planets range from about 3300 to 5500 kg m^{-3} , suggesting that the Moon is predominantly composed of rocky material, whereas Mercury mostly consists of metal (**Figure 2**). The Mercury-sized large icy satellites Ganymede, Callisto, and Titan have densities of only about 2000 kg m^{-3} . This implies that ice and rock/metal are present in about equal amounts in their interiors. Enceladus, Triton, Pluto, and Charon are icy bodies with similar ice-rock compositions ([Hussmann et al. 2006](#)).

The axial moment of inertia of a planet or satellite C can be inferred from the precession of its rotation axis in the presence of external torques exerted on its equatorial bulge. Measurement of the precession rate might require the placement of one or several landers on the surface of the body. For example, the combined analysis of Mars Global Surveyor tracking and Mars Pathfinder and Viking Lander range and Doppler data resulted in an improved value of the moment of inertia (MoI) factor of Mars, that is, the axial moment of inertia

normalized to the planet's mass times its radius squared, $C/MR^2 = 0.3654 \pm 0.0008$ ([Konopliv et al., 2006](#)). This value is much smaller than the MoI factor of 0.4 of a homogeneous sphere and is consistent with a significant concentration of mass toward the center due to a metallic core with a radius of about half that of the planet ([Schubert and Spohn, 1990; Sohl and Spohn, 1997; Sohl et al., 2005](#)). The axial moment of inertia of a planetary body can also be deduced from the figure of its gravitational field if the body is in hydrostatic equilibrium. However, this is a demanding requirement that is satisfied by only a few terrestrial-type bodies (see below).

10.02.2.2 Rotation and Tides

The internal mass distribution of a planetary interior can be inferred from its rotational and tidal response to external torques exerted on the planet or satellite (*see Chapter 10.04*). The torque applied to the body by the gravitational attraction of the Sun or a planet's satellite equals the scalar product of the body's inertia tensor and spin vector. Mercury and Venus have no satellites to affect their rotational states. The combined effect of the torques exerted by the tiny Martian moons, Phobos and Deimos, is much less than the torque exerted on Mars by the Sun ([Van Hoolst et al., 2003](#)). The torque applied to the Earth by the Moon decelerates the Earth's rotation due to tidal friction in the oceans and, to a minor extent, in the solid Earth ([Platzman, 1984](#)). In turn, the Moon is accelerated in its orbit and withdraws from Earth at a rate of a few centimeters per year. The rotation axes of Earth and Mars are tilted about axes normal to their orbital planes by 23.45° and 25.2° , respectively, thereby causing their spin axes to precess about the orbit normals under the influence of external gravitational torques.

Astronomical observations have revealed the precession rate of the Earth's spin axis and in turn the Earth's moment of inertia. Radio tracking of the Viking and Mars Pathfinder Landers and Mars Global Surveyor and Mars Odyssey orbiters has yielded similar information for Mars ([Esposito et al., 1992; Folkner et al., 1997; Yoder and Standish, 1997; Yoder et al., 2003; Konopliv et al., 2006](#)). Laser ranging observations of the Moon, made possible by reflectors placed on the lunar surface during the Apollo program, have determined the Moon's rotational state ([Dickey et al., 1994](#)). Because of its orbital eccentricity and ellipsoidal shape, the Moon is subject to periodic changes in angular acceleration in response

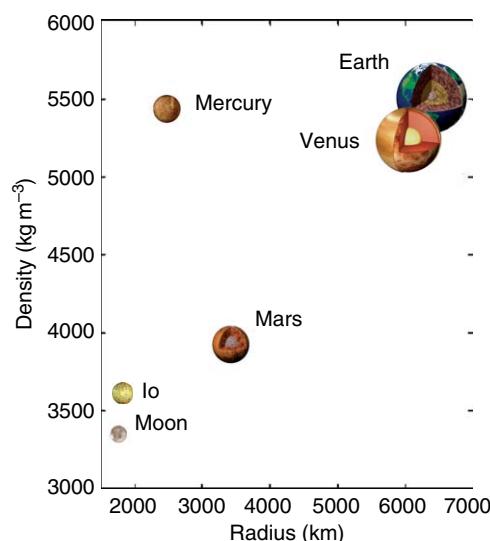


Figure 2 Radius–density relation of the terrestrial planets and the Moon. Note the unusually high mean density of Mercury implying that the planet's interior is predominantly composed of heavy elements such as iron.

to gravitational torques exerted by the Sun and the Earth. The amplitude of the related 27-day forced libration in longitude has been used to infer the moment of inertia of the Moon and the probable molten state of its central region (Yoder, 1981; Williams *et al.*, 2001; Khan *et al.*, 2004). In a similar way, measurements of Mercury's 88-day forced libration amplitude in longitude are expected to provide clues on the planet's internal mass distribution and coupling between core and mantle (Peale, 1976, 1988; Jehn *et al.*, 2004). It is unlikely, however, that the axial moment of inertia of Venus could be derived from observations of its rotational state alone since the planet's retrograde rotation is extremely slow and its rotation axis is more or less perpendicularly aligned to the plane of its nearly circular orbit (Yoder, 1997).

10.02.2.3 Gravity and Topography

Gravimetric and magnetic methods utilize measurements of potential fields in the vicinity of planetary bodies. The long-wavelength part of the external gravitational field provides information on the structure of planetary interiors. Each element of mass in a planet or satellite contributes to the external gravitational field according to Newton's law of gravitation. The gravitational force at a point exterior to the planet can be derived from the gradient or directional derivative of the gravitational potential caused by the total contribution of all mass elements when integrated over the entire volume of the body. In recognition of the first gravity experiments conducted by Galileo Galilei who measured fall times of objects possibly dropped from the Leaning Tower of Pisa, gravitational acceleration is given in units of $1 \text{ mGal} = 10^{-5} \text{ m s}^{-2}$.

The gravitational fields of planetary bodies are measured from orbiting spacecraft. The technique uses the Doppler effect on radio signals transmitted from the Earth to the spacecraft and back. The frequency shift of the radio signals returned by the spacecraft relative to the signal emitted from the ground station is proportional to the velocity component along the direction of vision or line of sight (LOS) and allows the calculation of the LOS acceleration of the spacecraft. These data can be used to calculate the gravitational field at a resolution determined by the altitude and the frequency of signal transmission to the spacecraft. Laser altimetry from orbiting spacecraft has been used to determine the topography of the Moon and Mars with high

accuracy. Radar altimetry data obtained by the Magellan spacecraft have provided topographic maps of Venus. The correlation of topography with the distribution of the Bouguer gravity (gravity corrected for contributions of topographic masses above a well-defined reference level) can be used to infer the degree of isostatic compensation of topography and the depth at which stresses are fully compensated due to rheological and/or compositional changes. Gravity anomalies provide important constraints on elastic lithosphere thickness, crust thickness, crust density, and load density and can be interpreted in terms of chemically or thermally induced lateral and depth variations of density. However, these interpretations suffer from the nonuniqueness inherent in any inversion of the gravity field (*see Chapter 10.05*).

The external gravitational field of a planet or satellite is generally described by a spherical harmonic representation of the gravitational potential as a function of latitude, longitude, and radial distance to the center of mass (CoM). The frequency content of the amplitude spectrum of the gravitational potential gives insight into the rheology and the mechanical and thermal properties of the lithosphere, the outer rigid shell of a planet or satellite. The leading term in the spherical harmonic expansion of the external gravitational potential is the ratio between the planetocentric constant GM and radial distance r representing the gravitational potential of the mass of the body concentrated in a point at its CoM. The next terms of the spherical harmonic representation are inversely proportional to r^3 and involve higher moments of the gravitational field caused by ellipsoidal distortions of the internal mass distribution. The magnitudes of these distortions are related to coefficients $J_2 = -C_{2,0}$ and $C_{2,2}$ for polar oblateness and equatorial ellipticity of the gravitational field, respectively. When the body is in hydrostatic equilibrium, its mean moment of inertia can be deduced from the knowledge of these coefficients with important implications for the internal structure. However, if the lithosphere is thick and strong enough to support stresses associated with surface or subsurface mass anomalies, the planet or satellite may substantially deviate from hydrostatic equilibrium. Furthermore, mantle density anomalies and core-mantle boundary (CMB) undulations may contribute to deviations from hydrostatic equilibrium. Accordingly, the Moon and Mercury are not expected to be fully hydrostatically compensated. The Earth is reasonably close to this state, but Mars is not, due to the uncompensated portion of the

Tharsis uplift. Though the gravitational field of Venus is known quite well, it is not known if the planet is in hydrostatic equilibrium. Therefore, the gravitational field cannot be used to give a reliable value of the planet's moment of inertia. A number of solid bodies in the outer solar system including the large icy satellites of Jupiter and Saturn, Ganymede, Callisto, and Titan, and the strongly tidally heated Jovian satellite Io are likely to be in hydrostatic equilibrium (*see* Chapter 10.15). The quadrupole gravitational fields of these Jovian satellites have been used to infer their moments of inertia (Schubert *et al.*, 2004).

10.02.2.4 Magnetic Fields

Magnetic field observations provide constraints on the interior structure of planets and moons (*see* Chapter 10.07). The existence of an intrinsic planetary magnetic field on a global scale is conclusive evidence that the body has a highly electrically conducting fluid region in its interior within which dynamo action creates the magnetic field. In the case of terrestrial planets the site of magnetic field generation is a metallic core that is at least partially liquid. Earth is the only terrestrial planet with a magnetic field that is unambiguously generated in its core. The nature of Mercury's magnetic field, that is, its mode of origin, is still uncertain though there is no question that Mercury has a large iron core (Schubert *et al.*, 1988). Jupiter's icy satellite Ganymede is the only other terrestrial-like solar system body whose magnetic field is generated by an active liquid metallic core dynamo. The discovery of Ganymede's magnetic field provided crucial evidence that the satellite is differentiated and has a metallic core (Schubert *et al.*, 1996). The absence of a planetary magnetic field, as in the cases of Venus and Mars, only implies the non-existence of an active dynamo: it does not imply the nonexistence of a metallic core nor does it require a metallic core to be solid.

An external planetary magnetic field can be represented by a spherical harmonic expansion of its magnetic potential in a manner similar to the representation of the gravitational field. It is possible to identify the core dynamo field in the spectrum of the magnetic field spherical harmonic coefficients and thereby determine the radius of the core. This procedure has been used by Voorhies *et al.* (2002) to estimate the radius of Earth's core as 3512 ± 64 km, in good agreement with the seismological radius (*see also* Elphic and Russell, 1978). A similar approach might lead to the determination of the radius of

Mercury's core once future spacecraft have mapped the complete spatial dependence of the planet's magnetic field, assuming that the field originates in a core dynamo.

Though not in possession of global internally generated magnetic fields at present, the Moon (Halekas *et al.*, 2001; Hood *et al.*, 2001; Lin *et al.*, 1998; Russell *et al.*, 1975, 1973) and Mars (Acuña *et al.*, 1999, 1998; Connerney *et al.*, 2004) have localized magnetic fields associated with crustal remanent magnetization. On the assumption that the crustal magnetization was acquired when these bodies had active dynamos in the past, even the crustal magnetization is evidence for the existence of a metallic core in these bodies. This is particularly important for the Moon which has a small metallic core that is difficult to detect by any method. Alternatively, Hood and Huang (1991) have proposed large-scale magnetization in lunar-basin forming impacts caused by plasma-induced antipodal amplification of ambient magnetic fields to explain the correlation of the largest magnetic field strengths with the antipodes of the largest lunar impact basins. The spatial variation of crustal magnetization and its associated magnetic fields also provide information on the internal structure of a body at shallow depth and the internal and surface processes that have affected the crust.

10.02.2.5 Electromagnetics

The phenomenon of electromagnetic induction can be utilized to probe planetary interiors. The method takes advantage of the electrically conducting nature of planetary materials and the time variability of the magnetic fields experienced by planetary bodies. If an electrically conducting object experiences a time-variable magnetic field, electrical currents are induced in the object (Faraday's law of induction). These induced currents in turn generate a magnetic field. If the inducing and induced magnetic fields can be measured, then the electrical conductivity of the object can be inferred. This method has been successfully used to infer the electrical conductivity of the Earth (Banks, 1969; Chapman and Price, 1930; Hobbs, 1987; Lahiri and Price, 1939; Rikitake, 1966; Roberts, 1986; Schuster, 1889; Tarits, 1994) and the Moon (Schubert and Schwartz, 1969; Sonett, 1982; Sonett *et al.*, 1971). It has also been used to find subsurface liquid salt water oceans on the Galilean satellites Europa, Ganymede, and Callisto (Khurana *et al.* 1998; Kivelson *et al.*, 1999, 2000, 2002, 2004; Zimmer *et al.* 2000). In the case of the

Earth, time variations in the external magnetic field are forced by variable ionospheric and magnetospheric current systems. The Moon experiences the variable magnetic field of the solar wind when it is outside the Earth's magnetosphere. The Galilean satellites experience the variability of the rotating, tilted Jovian magnetic field. On Earth, the external forcing magnetic field can be separated from the internal induced magnetic field by a network of surface magnetometers. More recently, satellite data have also been used for this purpose (Olsen 1999; Constable and Constable 2004). For the Moon, magnetometers on the surface and an orbiting spacecraft measured the solar wind forcing field and inductive response during the era of lunar exploration by Apollo spacecraft in the late 1960s and early 1970s. A magnetometer on the Galileo spacecraft measured the induced magnetic fields of Europa, Ganymede, and Callisto. Knowledge of electrical conductivity inside a planetary body indirectly constrains the composition, temperature, and volatile content of the body through the dependence of the conductivity on these properties (Hood, 1986; Hood *et al.*, 1982; Hood and Jones, 1987; Hood and Sonett, 1982; Hood and Zuber, 2000; Sonett *et al.*, 1972). The electromagnetic sounding approach involves two inversion steps, inference of the electrical conductivity from magnetic measurements and inference of composition, temperature, and volatile content from the conductivity.

Magnetic field measurements obtained during the passage of the Moon through the Earth's geomagnetic tail provide a means of detecting the lunar core (Hood *et al.*, 1999). For several days each month the Moon passes through the north and south lobes of the geomagnetic tail and experiences a vacuum-like environment with a near-steady uniform magnetic field. This results in an induced lunar magnetic dipole that acts to exclude the tail lobe field from the Moon's interior. The induced magnetic dipole is produced by electric currents that flow near the surface of the highly electrically conducting metallic core (Hood *et al.*, 1999). The lunar-induced magnetic dipole has been detected by the magnetometer on the Lunar Prospector spacecraft; the measured value of the induced dipole moment implies a core of radius 340 ± 90 km (Hood *et al.*, 1999).

10.02.2.6 Seismology

The propagation of seismic waves following quakes or impacts can be used to infer the structure of planetary interiors. Lognonné (2005) and Lognonné

and Mosser (1993) have provided reviews on planetary seismology (see Chapter 10.03). Whereas surface waves are confined to near-surface layers, body waves travel through the interior. The first body waves arriving at a seismic station are longitudinal or P waves that involve material compression and rarefaction. The second are transverse or S waves involving shear motion perpendicular to the direction of propagation. Body waves are subject to reflection, transmission, and conversion from P to S waves and vice versa at internal boundaries at which prominent density and seismic velocity changes occur. A network of at least four seismic stations is required to infer the source location from a comparison of the seismic station records. Free oscillations on a global scale can be excited if the magnitudes of seismic events are sufficiently large. For the Earth, seismology has not only revealed the existence of a metallic core surrounded by a rocky mantle, but it has shown that the core consists of a solid inner core composed of iron and nickel and a liquid outer shell. The analysis of surface waves has provided important clues on the chemical layering of the Earth's crust. Though seismic stations were placed on the Moon's surface during the Apollo exploration era, the collected data are unable to confirm the presence or absence of a small metallic core and do not allow construction of detailed models of the lunar crust and mantle. Seismic instruments placed on the surface of Mars by the Viking mission did not provide data that could be used to determine the planet's internal structure. However, the deployment of seismometers will be central to any future network mission to Mars that attempts to determine the planet's gross interior structure (Lognonné *et al.*, 2000; Dehant *et al.*, 2000). Tidally induced gravity variations raised by Phobos may also permit probing of the deep Martian interior (Van Hoolst *et al.*, 2003).

Promising target bodies for seismic exploration in the outer solar system are the Jovian satellite Europa and the Saturnian satellite Titan due to the likely presence of satellite-wide internal liquid layers beneath the outer ice shells (Grasset *et al.*, 2000; Sohl *et al.*, 2003; Tobie *et al.*, 2005). A passive seismic experiment placed on the surface of Europa would be able to detect seismic activity generated by the formation of cracks, tidally induced quakes, or natural impacts. The thickness of the ice shell could be obtained from low-frequency band observations (0.1–10 Hz) of trapped surface waves (Kovach and Chyba, 2001; Lee *et al.*, 2003).

10.02.2.7 Surface Geology and Composition

The geology and composition of a planet's surface provide important clues into what lies inside the body. These glimpses into the interior, though generally qualitative in nature, are valuable nonetheless and complement more quantitative constraints on internal structure. The basaltic crust of Mercury and the basaltic lowlands of Venus are evidence of the internal differentiation of these bodies. The anorthositic lunar highlands not only attest to the differentiation of the Moon, but they argue for a global-scale magma ocean early in lunar history (Wood *et al.*, 1970; Smith *et al.*, 1970; Warren, 1985). Mercury's high density suggests it is iron rich, yet spectral evidence suggests that its surface is poor in oxidized iron (Vilas, 1985; Blewett *et al.*, 1997), consistent with the segregation of its iron into a central core following differentiation under reducing conditions (Sprague *et al.*, 1994; Schubert *et al.*, 1988).

The crust dichotomy of Mars represents an ancient feature of early Noachian age ($>3.5\text{--}3.7\text{ Gyr}$ ago) that is preserved in the surface geology and tectonics, the cratering record, and the planet's gravity and magnetic field. The heavily cratered southern highlands crust on Mars and the resurfaced northern lowlands of the planet indicate early differentiation and internal dynamical activity (Watters and McGovern, 2005), though removal of the northern crust by one or multiple giant impacts about 4 Gyr ago has also been hypothesized (Wilhelms and Squyres, 1984). However, the detection of quasi-circular depressions in the northern lowlands, interpreted as buried impact craters and basins, suggests a similar age for the lowland and the ancient southern highland crust (Frey *et al.*, 2002). This is in favor of an endogenic origin of the crustal dichotomy involving dynamical processes like degree-1 mantle convection (Schubert and Lingenfelter, 1973; Lingenfelter and Schubert, 1973; Zhong and Zuber, 2001; Roberts and Zhong, 2006), thin-crust formation by early plate tectonics, and the solidification and convective instability of a pristine magma ocean on Mars (Elkins-Tanton *et al.*, 2003b). The ancient dichotomy boundary was modified then by fluvial, aeolian, and glacial processes, as well as by widespread volcanic and sedimentary resurfacing of the northern lowlands (Watters and McGovern, 2005).

Surface geology and signs of endogenic activity are particularly important for inferring the internal structure of outer solar system moons. The heavily

cratered surfaces of Jupiter's icy moon Callisto and Saturn's icy moon Rhea support the inference that Rhea is essentially undifferentiated while Callisto is partially differentiated (ice from rock) (Schubert *et al.*, 2004; Anderson and Schubert, 2007). In contrast, the highly modified surfaces of the Jovian satellites Ganymede and Europa and the Saturnian satellite Enceladus are consistent with the separation of ice and rock in their interiors (Schubert *et al.*, 2004; Porco *et al.*, 2006; Schubert *et al.*, in press). The ubiquitous and ongoing volcanic activity of Jupiter's satellite Io leaves little doubt that the moon has differentiated into an iron-rich core and silicate mantle (Moore *et al.*, 2006). Thermal anomalies detected in Enceladus' south polar region and the active plumes spewing water from beneath the south polar surface of Enceladus (Porco *et al.*, 2006; Spencer *et al.*, 2006) provide additional persuasive evidence for the formation of a rock core and ice mantle inside this body.

While not definitive in themselves, surface geologic and compositional data help to constrain models of planetary and satellite interiors, especially when more quantitative observations are limited.

10.02.2.8 Material Properties

There are only a few kimberlite rock samples brought to the Earth's surface by violent volcanic eruptions that directly probe the deep interior of a terrestrial planet. Diamond deposits occasionally found in kimberlite rocks suggest that confining pressure must have exceeded the pressure at which diamonds are stable. Meteorites originally released from the surfaces of the Moon and Mars during one or several giant impact events represent another important data source on the composition of terrestrial-planet interiors. Much of our knowledge about the interior structure and evolution of the Earth and other terrestrial planets, however, comes from high-pressure mineral physics that includes laboratory experiments and computational studies. Bass (2004) summarizes current and future research activities in this highly interdisciplinary and rapidly evolving field of research (see Volume 2).

Earth and planetary material properties are affected by variable compositions and thermodynamic conditions. The core sulfur content may range from close to eutectic to iron-rich compositions with important implications for the physical state and density stratification of planetary cores (Fei *et al.*, 1995, 1997; Sanloup *et al.*, 2000; Kavner

et al., 2001; Balog *et al.*, 2003) due to the substantial melting point reduction with increasing sulfur concentration (Boehler, 1992, 1996a). Mantle rheology is affected by predominant mineral phase assemblages, water content, and typical mineral grain-size distribution. It is possible that mantle mineralogies of other terrestrial bodies are not dominated by olivine but rather by mixtures of pyroxene minerals. Experimental and theoretical studies of two-phase rheologies (e.g., rock–volatile, rock–melt) are particularly useful for a better understanding of the influence of volatiles like water and carbon and solid gas hydrates on deeply situated geodynamic processes (Sloan, 2003), like the subduction of lithospheric plates on Earth, the rheological behavior of crust and mantle rocks, the kinetics of pressure-induced mineral phase transformations, and the extraction of partial mantle melts (Bass, 2004).

Nowadays, sophisticated tools are available to study the properties of matter under extreme pressure and temperature conditions that prevail in planetary interiors (Liebermann, 2005). From shock and static compression experiments using diamond-anvil facilities, densities, EOS parameters (Hemley and Ashcroft, 1998; Hemley, 2006), and elastic properties of solid mineral phase assemblages (Anderson *et al.*, 1992; Bina and Helffrich, 1992), as well as high-pressure properties of melts (Boehler, 1996b, 1996c) are derived. In these experiments, relatively small samples are subjected to conditions that may prevail deep in the Earth's mantle or even in its metallic core ($P > 135$ GPa, $T > 3000$ K). Future experiments are expected to provide important information on high-pressure transport properties like thermal conductivity and kinematic viscosity, the kinetics of chemical reactions at the CMB, and mineral-phase equilibria up to several tens of GPa. While inelastic X-ray diffraction methods are used to measure sound velocities at very high pressures and temperatures, neutron scattering observations are well-suited to determine the structural properties of silicate melts and aqueous solutions (Bass, 2004). Laboratory experiments need to be augmented by high-performance computations, for example, to simulate diffusion-limited transport processes occurring at atomic scales. These simulations are important to address the influence of oxygen fugacity on mantle rock phase transformation kinetics and the element solubility of light-element admixtures in liquid iron alloys (Badding *et al.*, 1991; Wood, 1993).

10.02.3 Interior Structure and Composition

In the following, we consider two- and three-layer density models before we discuss radially symmetric, depth-dependent models of interior structure.

10.02.3.1 Two- and Three-Layer Structural Models

Models of the internal density distribution of terrestrial planets suffer from an inherent nonuniqueness since there are usually fewer constraints than unknowns. These models are required to satisfy two constraints, the mean density ρ as derived from the total radius R and mass M_p and the mean moment of inertia I that can be determined from the quadrupole moments of the gravitational field and the frequency of precession of the rotation axis.

Given the paucity of information available, three-layer models representing the core, the mantle, and the crust can be considered as useful approximation to the structure of terrestrial-planet interiors. The corresponding structural equations are given by

$$\rho = \rho_s + (\rho_c - \rho_m) \left(\frac{R_c}{R} \right)^3 + (\rho_m - \rho_s) \left(\frac{R_m}{R} \right)^3 \quad [1]$$

$$\frac{I}{MR^2} = \frac{2}{5} \left(\frac{\rho_s}{\rho} + \frac{\rho_c - \rho_m}{\rho} \left(\frac{R_c}{R} \right)^5 + \frac{\rho_m - \rho_s}{\rho} \left(\frac{R_m}{R} \right)^5 \right) \quad [2]$$

where the core radius R_c , the crust–mantle radius R_m , the crust density ρ_s , the mantle density ρ_m , and the core density ρ_c are unknown. Even two-layer interior structure models lacking a crust layer ($\rho_s = 0$, $R_m = R$) would have fewer constraints than unknowns, that is, ρ_c , R_c , and ρ_m .

The mean density of a two-layer spherical body is given by

$$\rho R^3 = (\rho_c - \rho_m) R_c^3 + \rho_m R^3 \quad [3]$$

from which the relative core radius R_c/R , core mass fraction M_c/M_p , and dimensionless mean MoI factor are obtained according to

$$\frac{R_c}{R} = \left(\frac{\rho - \rho_m}{\rho_c - \rho_m} \right)^{1/3} \quad [4]$$

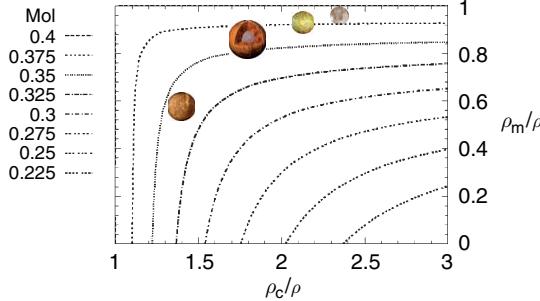


Figure 3 Contours of the mean moment-of-inertia factor MoI for two-layer structural models of planetary interiors as a function of core and mantle density ρ_c and ρ_m , respectively, relative to mean density ρ . The symbols shown in the diagram represent putative locations of the Moon, Io, Mars, and Mercury.

$$\frac{M_c}{M_p} = \frac{\rho_c}{\rho} \left(\frac{R_c}{R_p} \right)^3 = \frac{\rho_c (\rho - \rho_m)}{\rho (\rho_c - \rho_m)} \quad [5]$$

and

$$\text{MoI} = \frac{I}{M_p R^2} = \frac{2}{5} \left(\frac{(\rho - \rho_m)^{5/3}}{\rho(\rho_c - \rho_m)^{2/3}} + \frac{\rho_m}{\rho} \right) \quad [6]$$

respectively. If hydrostatic equilibrium is assumed, the corresponding increase of density with depth is equivalent to the requirement $\rho_c \geq \rho$ and $0 \leq \rho_m \leq \rho$. In Figure 3, contours of MoI are shown for plausible ranges of core and mantle densities normalized to the mean density, ρ_c/ρ and ρ_m/ρ , respectively.

10.02.3.2 Multilayer Structural Models

A commonly used approach to modeling the interior structure of a terrestrial planet given its mass, radius, and moment of inertia (see Wood *et al.* (1981) for a review) converts the estimated bulk composition into a mineralogical model and adopts potential temperatures for the mantle layers and the core from thermal history calculations. The potential temperature is the temperature extrapolated adiabatically to the surface pressure. The radial density distribution is then calculated under the assumption of hydrostatic and thermal equilibrium using an EOS to correct for compression and thermal expansion. The total mass and the MoI are calculated from the model and compared with the data. This approach can be modified by simultaneously calculating the thermal and mechanical structure of a terrestrial body (Sohl and Spohn, 1997). In order to derive a self-consistent structural model, STP values of the density, the

bulk modulus, and the thermal expansivity are calculated from laboratory data for the individual layers of the model.

10.02.3.2.1 Governing equations

If a spherically symmetric planet in perfect mechanical and thermal equilibrium is assumed, the following set of differential equations for mass m , iron mass m_{Fe} , mean moment of inertia θ , acceleration of gravity g , pressure p , and heat flux q can be derived from fundamental principles:

$$\frac{dm}{dr} = 4\pi r^2 \rho_r \quad [7]$$

$$\frac{dm_{Fe}}{dr} = x_{Fe} \frac{dm}{dr} \quad [8]$$

$$\frac{d\theta}{dr} = \frac{8}{3}\pi r^4 \rho_r \quad [9]$$

$$\frac{dg}{dr} = 4\pi G \rho_r - 2 \frac{g}{r} \quad [10]$$

$$\frac{dp}{dr} = -\rho_r g \quad [11]$$

$$\frac{dq}{dr} = \rho_r \varepsilon_r - 2 \frac{q}{r} \quad [12]$$

where r is the radial distance from the center of the planet, G is the gravitational constant, ρ is the density, x_{Fe} is the concentration of iron per unit mass, and ε is the specific heat production rate. The subscript r indicates quantities that are local functions of p , T , and composition.

The mean MoI factor $I/M_p r_p^2$ should be used for the construction of spherically symmetric models and can be derived from the planet's observed polar MoI factor $C/M_p r_p^2$ and the second-degree coefficients of the spherical harmonic representation of its gravitational field, \mathcal{J}_2 and $\mathcal{J}_{2,2}$. The mean and polar MoI factors are related to each other by

$$\frac{I}{M_p r_p^2} = \frac{C}{M_p r_p^2} - \frac{2}{3} \mathcal{J}_2 = \frac{C}{M_p r_p^2} + \frac{2}{3} C_{2,0} \quad [13]$$

where

$$\mathcal{J}_2 = \frac{1}{M_p r_p^2} \left[C - \frac{A + B}{2} \right] = -C_{2,0} \quad [14]$$

is the planet's gravitational oblateness and $A < B < C$ are the planet's principal equatorial and polar moments of inertia, respectively. A more comprehensive account including minor nonhydrostatic corrections to the mean MoI factor due to symmetric distributions of topographical masses like, for

example, the Tharsis rise on Mars is provided by Sohl *et al.* (2005).

The base of the rheological lithosphere is defined with the help of the temperature T_{rheo} at which subsolidus creep becomes effective over geologic timescales and which is taken as 0.6 times the solidus temperature T_m (Meissner and Vetter 1979) or about 800°C. The rheological lithosphere is the upper part of the thicker thermal lithosphere which additionally comprises the upper cold thermal boundary layer of the convecting mantle and thus represents the thermally conductive outermost layer of the planet (Spohn, 1991). The boundary temperature at the base of the thermal lithosphere T_{lith} is taken to be 0.85 T_m (Ranalli 1987) or about 1400°C. There is an additional core–mantle thermal boundary layer at the bottom of the mantle. The thickness δ of the CMB layer is calculated from the local critical Rayleigh number for marginal stability of the layer. The local critical Rayleigh number is (Jarvis and Peltier, 1989)

$$Ra_c = \frac{g\alpha}{\kappa\nu k} q_b \delta^4$$

where κ is the thermal diffusivity, α is the thermal expansivity, and the kinematic viscosity ν is evaluated at the geometrically averaged temperature of the boundary layer; q_b is the sum of the basal heat flux and the heat production rate per unit area $\rho\varepsilon\delta$ and k is the thermal conductivity. Within the boundary layers, heat is transported radially by conduction and the temperature T satisfies

$$\frac{dT}{dr} = -\frac{q}{k_r} \quad [15]$$

Between the boundary layers energy is primarily carried by convection and the temperature gradient is assumed to be approximately equal to the adiabatic temperature gradient (Stacey, 1977)

$$\frac{dT}{dr} = T \frac{\gamma_r}{K_{S,r}} \frac{dp}{dr} \quad [16]$$

where $\gamma = \alpha K_S / \rho c_p$ is the thermodynamic Grüneisen parameter, and K_S is the adiabatic bulk modulus.

The radial positions of pressure-induced mantle phase boundaries like the exothermic olivine– β -spinel and β -spinel– γ -spinel transitions and the endothermic γ -spinel–perovskite transition, can be obtained from the intersections of the temperature profile calculated from eqns. [12] through [16] and the specific Clausius–Clapeyron curves.

The set of basic differential equations [7]–[16] can be separated into two subsets that are coupled through the density ρ . The mechanical properties of the interior are calculated from eqns [7]–[11], while eqns. [12]–[16] give the thermal structure of the model. Because ρ depends more strongly on pressure than on temperature, a feature common to most planetary bodies, except the strongly tidally heated Jovian satellite Io and perhaps the Earth's Moon, it is possible to calculate the density distribution by using an isothermal fourth-order Eulerian finite strain–Murnaghan–Birch EOS and by applying temperature corrections through a calculation of the thermal pressure as described further below (Stacey *et al.*, 1981) and illustrated in Figure 4.

10.02.3.2.2 Equation of state

STP parameter values of density ρ_{ref} , mean atomic mass $\bar{\mu}$, thermal expansivity α_{ref} , isothermal bulk modulus $K_{T,\text{ref}}$, and rigidity G_{ref} of an assemblage of n components can be calculated from

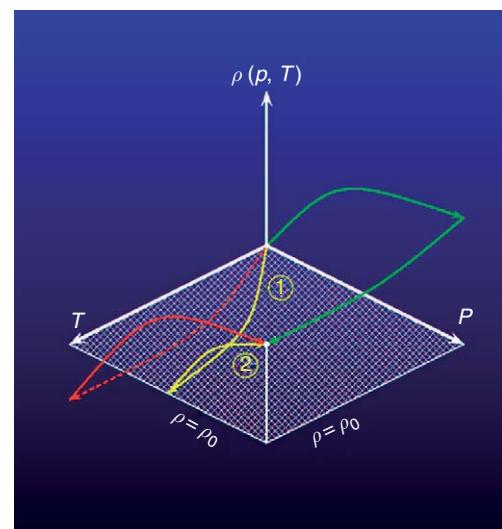


Figure 4 Schematic view of the extrapolation of density $\rho(p, T)$ to elevated pressures p and temperatures T . The origin of the coordinate system refers to an arbitrarily chosen reference density $\rho_0(p_0, T_0)$ at fixed (standard) pressure and temperature conditions p_0 and T_0 , respectively. First the increase of thermal pressure caused by thermal expansion is calculated (path 1), then the isothermal Birch–Murnaghan equation (path 2) is applied to account for pressure-induced compression. Also shown are alternate thermodynamic paths. Adapted from Wood JA, Anderson DL, Buck WR *et al.* (1981) Geophysical and cosmochemical constraints on properties of mantles of the terrestrial planets. In: Kaula WM (ed.) *Basaltic Volcanism on the Terrestrial Planets*, pp. 633–699. New York: Pergamon.

volume-weighted averages applying the widely used Voigt–Reuss–Hill method (Watt *et al.*, 1976):

$$\rho_{\text{ref}} = \left(\sum_{i=1}^n \frac{x_i}{\rho_i} \right)^{-1} \quad [17]$$

$$\bar{\mu} = \left(\sum_{i=1}^n \frac{x_i}{\mu_i} \right)^{-1} \quad [18]$$

$$\alpha_{\text{ref}} = \rho_{\text{ref}} \sum_{i=1}^n \frac{x_i \alpha_i}{\rho_i} \quad [19]$$

$$K_{T,\text{ref}} = \left(\rho_{\text{ref}} \sum_{i=1}^n \frac{x_i}{\rho_i K_{T,i}} \right)^{-1} \quad [20]$$

$$G_{\text{ref}} = \left(\rho_{\text{ref}} \sum_{i=1}^n \frac{x_i}{\rho_i G_i} \right)^{-1} \quad [21]$$

where x_i is the mass fraction of an individual crust, mantle, or core component with STP density ρ_i , mean atomic mass $\bar{\mu}_i$, thermal expansivity α_i , isothermal bulk modulus $K_{T,i}$ and rigidity G_i . This is the simplest bounding method that is preferentially applied to isotropic composites by calculating arithmetic means of bounds in which stresses (Reuss bound) and strains (Voigt bound) are assumed to be uniform. For two-phase composites and polycrystals, however, Hashin–Shtrikman bounds are preferred over the Voigt–Reuss–Hill average (Watt *et al.*, 1976).

Since internal temperatures of the terrestrial-type bodies exceed the Debye temperature, the quasi-harmonic approximation is valid and the product of thermal expansivity α and isothermal incompressibility K_T

$$\alpha K_T = - \left(\frac{1}{\rho} \frac{\partial \rho}{\partial T} \right)_p \times \left(\rho \frac{\partial p}{\partial \rho} \right)_T = \left(\frac{\partial p}{\partial T} \right)_V \quad [22]$$

can be considered to be constant throughout chemically homogeneous layers (Anderson *et al.*, 1992). Under these circumstances, the thermal pressure (cf. Figure 4, path 1)) is given by

$$p_{\text{th}} = \int_{T_{\text{ref}}}^T \left(\frac{\partial p}{\partial T} \right)_V dT = \int_{T_{\text{ref}}}^T \alpha K_T dT \approx \alpha_{\text{ref}} K_{T,\text{ref}} (T - T_{\text{ref}}) \quad [23]$$

and can be used to include the temperature effect in the EOS parameters (Anderson, 1984).

From the thermodynamic identity (Anderson, 1989)

$$\left(\frac{dM}{dT} \right)_V = \left(\frac{dM}{dT} \right)_p + \alpha K_T \left(\frac{dM}{dp} \right)_T \quad [24]$$

for any thermodynamic function M and Swenson's law (Anderson *et al.*, 1992)

$$\left(\frac{dK_T}{dT} \right)_V = 0 \quad [25]$$

the first temperature derivatives $(dK_T/dT)_p$, $(d\alpha/dT)_p$, and $(dG/dT)_p$ used at constant pressure in the high-temperature limit for making linear temperature corrections to the EOS parameters can be attributed to well-known physical properties under STP conditions according to

$$\left(\frac{dK_T}{dT} \right)_p = -\alpha_{\text{ref}} K_{T,\text{ref}} \left(\frac{dK_T}{dp} \right)_{\text{ref}} \quad [26]$$

$$\left(\frac{d\alpha}{dT} \right)_p = \alpha_{\text{ref}}^2 \left(\frac{dK_T}{dp} \right)_{\text{ref}} \quad [27]$$

$$\left(\frac{dG}{dT} \right)_p = -\alpha_{\text{ref}} G_{\text{ref}} \left(\frac{dK_T}{dp} \right)_{\text{ref}} \quad [28]$$

where $(dK_T/dp)_{\text{ref}}$ is the first pressure derivative of the isothermal bulk modulus as obtained from Voigt–Reuss–Hill averaging over individual components in the reference state. Furthermore, the first pressure derivative of rigidity is simply taken as (Poirier and Liebermann 1984)

$$\left(\frac{dG}{dp} \right)_{\text{ref}} = \frac{G_{\text{ref}}}{K_{T,\text{ref}}} \left(\frac{dK_T}{dp} \right)_{\text{ref}} \quad [29]$$

Using the truncation in Eulerian strain at fourth order, we subsequently infer the pressure dependence of the elastic moduli G and K_T and its local pressure derivative dK_T/dp (Stacey *et al.*, 1981; Bina and Helffrich, 1992):

$$K_T = K_{T,0} (1 - 2f)^{5/2} \left\{ 1 - f \left(5 - 3 \left(\frac{dK_T}{dp} \right)_{\text{ref}} \right) + \frac{f^2}{2} \left[9K_{T,0} K''_{T,0} + \left(3 \left(\frac{dK_T}{dp} \right)_{\text{ref}} - 7 \right) \times \left(3 \left(\frac{dK_T}{dp} \right)_{\text{ref}} - 5 \right) \right] \right\} \quad [30]$$

$$G = G_0 (1 - 2f)^{5/2} \left\{ 1 - f \left(5 - 3 \left(\frac{dG}{dp} \right)_{\text{ref}} \frac{K_{T,0}}{G_0} \right) + \frac{f^2}{2} \left[9G_0'' \frac{K_{T,0}^2}{G_0} + 9 \left(\frac{dG}{dp} \right)_{\text{ref}} \frac{K_{T,0}}{G_0} \times \left(\left(\frac{dK_T}{dp} \right)_{\text{ref}} - 4 \right) + 35 \right] \right\} \quad [31]$$

$$\frac{dK_T}{dp} = \left(\frac{dK_T}{dp} \right)_{\text{ref}} + 3K_{T,0}K''_{T,0}f \quad [32]$$

with

$$f = \frac{1}{2} \left[\left(\frac{\rho}{\rho_0} \right)^{2/3} - 1 \right] \quad [33]$$

where the double prime indicates the second derivative with respect to pressure. Since reliable measurements of $K''_{T,0}$ and G''_0 are not readily available for terrestrial mantle rocks, one may assume

$$G''_0 \frac{K_{T,0}^2}{G_0} \approx -\frac{35}{9} \quad [34]$$

and (Hofmeister, 1991)

$$K_{T,0}K''_{T,0} \approx - \left(\frac{dK_T}{dp} \right)_{\text{ref}}^2 + 7 \left(\frac{dK_T}{dp} \right)_{\text{ref}} - \frac{143}{9} \quad [35]$$

Consequently, the pressure effect on density is attributed to isothermal compression (cf. Figure 4, path 2) and, finally, eqns [30] and [33] are iteratively solved for local density ρ .

To calculate the seismic structure of the interior in terms of the P- and S-wave velocities

$$v_p = \sqrt{\frac{K_S + (4/3)}{\rho}} \quad [36]$$

and

$$v_s = \sqrt{\frac{G}{\rho}} \quad [37]$$

the isothermal–adiabatic transformation

$$\frac{K_S}{K_T} = 1 + \gamma \alpha T \quad [38]$$

between adiabatic bulk modulus K_S and its isothermal counterpart K_T is employed by using Slater's formulation

$$\gamma \approx \frac{1}{2} \frac{dK_T}{dp} - \frac{1}{6} \quad [39]$$

as an approximation of the thermal Grüneisen parameter γ (Stacey, 1977).

10.02.3.2.3 Boundary conditions

The set of basic differential equations [7]–[16] can be solved by numerical integration with respect to the following boundary conditions. The central boundary conditions at $r=0$ are

$$\begin{aligned} m &= 0 \\ m_{\text{Fe}} &= 0 \\ \theta &= 0 \\ g &= 0 \\ p &= p_c \\ q &= 0 \\ T &= T_c \end{aligned} \quad [40]$$

Since there are three observational constraints on the model, the mass, the radius, and the polar moment of inertia or the global Fe/Si ratio, respectively, we can choose three parameters as adjustable, the values of which are iteratively adjusted such that the observational constraints can be satisfied. These parameters are the central pressure p_c , the central temperature T_c , and the pressure at the core–mantle boundary p_{cmb} .

The surface boundary conditions at $r=r_p$ are

$$\begin{aligned} m &= M_p \\ m_{\text{Fe}} &= M_{\text{Fe}} \\ \theta &= I \\ g &= g_p \\ p &= p_p \\ q &= q_p \\ T &= T_p \end{aligned} \quad [41]$$

where the mean moment of inertia I can be approximately given by the polar moment C . While the mass M_p and the mean surface values of gravity g_p , pressure p_p and temperature T_p have been derived from spacecraft and Earth-based observations, the surface heat flux q_p values of almost all terrestrial bodies are unknown at present. Therefore, the boundary condition for the heat flux density is taken at the core–mantle boundary rather than at the surface. This is convenient because the heat flux in the mantle and in the crust depends on the bulk chemistry. The heat flow from the core into the base of the mantle q_{cmb} strongly depends on the thermal history of the core and can be assessed from thermal history calculations (see Chapter 10.09).

10.02.3.2.4 Numerical solution

A modified predictor–corrector integration scheme using the Adams–PECE method (Shampine and Gordon, 1975) is suitable to numerically solve eqns [7]–[16]. Furthermore, a shooting method using multidimensional globally convergent Newton–Raphson

iteration (Press *et al.*, 1992) can be employed in order to simultaneously adjust the free parameters of the system, for example, p_c , T_c , p_{cmb} , p_{cr} , q_p , and the global ratio Fe/Si or the MoI factor $C/M_p r_p^2$. Integration proceeds from the center outwards and, at the same time, from the surface inwards. The solutions are required to match at an arbitrarily chosen matching point within a prescribed numerical accuracy limit. If present, the olivine–spinel transition may serve as matching point with continuous iteration until the successful solution requiring continuity of all components is met at the internal fitting point within the numerical accuracy limit. For the sake of rapid convergence to the successful solution, the initial set of starting parameters is constrained to lie sufficiently close to the final set of parameters. An educated guess of the initial set can be obtained by using the above analytical solutions for a three-layer structural model having the planet’s mass, iron mass fraction, basaltic crust density, and homogeneously distributed crust/mantle heat sources.

10.02.4 Earth as a Type Example of a Terrestrial Planet

Inferences about the structure, composition, and mineralogy of the planets are guided by our knowledge of the Earth. In some cases, Venus for example, observational constraints on the interior are so minimal that the best we can do is to argue by analogy with Earth. Our knowledge of Earth’s interior, though partially based on the same types of observations available for the planets, rests largely on seismological data. The Moon is the only planetary body other than Earth for which we have seismological data and those data are limited in both quality and quantity. The structure, composition, and mineralogy of the Earth are discussed at length in other volumes of this treatise (see Volumes 1 and 2) and in the *Treatise of Geochemistry* (see Volumes 1–3.) The brief summary we present here is intended only to place the blurry pictures of planetary interiors in the context of the sharper picture of Earth’s interior.

10.02.4.1 General

The basic structure of the Earth is that of a three-layer sphere. The innermost layer is a metallic core of radius 3486 km. Surrounding the core is a rocky spherical shell or mantle of thickness 2885 km and encircling the mantle is a thin rocky crust distinct in

composition from the underlying mantle. The core is itself divided into two parts, an inner solid core of radius 1217 km surrounded by an outer liquid shell about 2269 km thick. It is believed that the inner core has formed as a consequence of the cooling of the Earth over geologic time, a process that has resulted in the partial solidification of the core from the inside out. The mantle is itself divided into two parts, the upper mantle about 660 km thick and the lower mantle about 2225 km thick. The mantle subdivisions are based on the occurrence of solid–solid phase transitions in the mantle rock, about which more is discussed below. The crust is neither uniform in composition or thickness. An approximately 6-km-thick crustal layer covers the floors of the oceans while a compositionally distinct crust of about 30 km thickness comprises the continents. The basic structure of the Earth is shown in Figure 5.

10.02.4.2 Interior Structure

The overall structure of the Earth is generally believed to have been set early in Earth’s evolution, within tens to about 100 Myr after accretion was complete (see Volume 9). The gravitational potential energy released upon accretion was large enough to melt the Earth’s interior yielding liquid metal, mostly iron, that sank toward the center and accumulated to form the metallic core. The process of core formation could have begun before accretion was complete. Core formation itself releases additional energy that

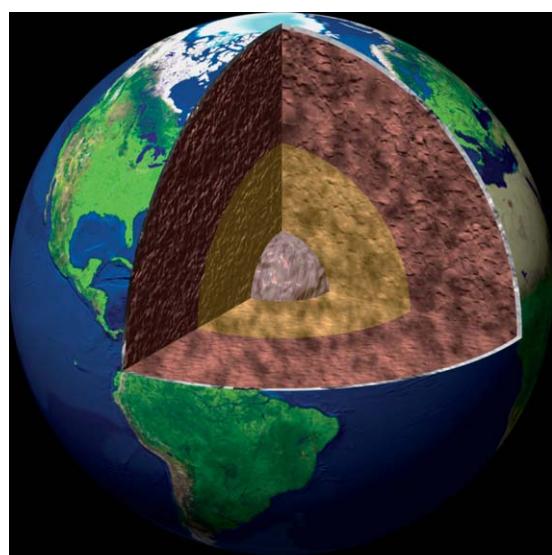


Figure 5 Cut-away view of the Earth’s interior. © Calvin J. Hamilton.

could have contributed to melting the mantle. Formation of the core could have been a runaway process. The timing of crustal formation is more uncertain. Certainly, oceanic crust (basalt) is being produced today by melting of mantle rocks at mid-ocean ridges. This process has occurred throughout Earth's history, as long as plate tectonics has been active. Basaltic volcanism does not require plate tectonics, as is evident from the surfaces of other planets, so basaltic crust was likely produced on Earth early in its evolution by plate tectonics or other processes. Oceanic crust is recycled back into the mantle by plate tectonics. The oldest oceanic crust on Earth is only about 200 Myr. The more difficult problem in crustal evolution is the history of the continental crust. The more siliceous continental crust is believed to have been produced by multiple stages of melting in the presence of water. The process is occurring in the andesitic volcanoes at convergent plate margins. The main question is how much continental crust existed on Earth as a function of time throughout Earth's history (see Chapter 9.07).

10.02.4.3 Composition

The bulk composition of the Earth, aside from its volatiles, is believed to be that of the meteorites classified as CI chondrites, the assumed primordial building blocks of the Earth. Accordingly, the Earth is 32.7 wt.% Fe, 15.4 wt.% Mg, 14.2 wt.% Si, 1.89 wt.% Ni, 1.71 wt.% Ca, and 1.59 wt.% Al (McDonough and Sun, 1995). Other elements are of course present in relatively smaller amounts. Much of the Earth's Fe and Ni has segregated into the core which contains about 87.5 wt.% Fe, 5.4 wt.% Ni, 0.95 wt.% Cr, 0.5 wt.% Mn, and unknown amounts of light elements such as O and S (McDonough and Sun, 1995). The light elements of the core are preferentially excluded from the inner core upon its solidification and concentrated in the liquid outer core. The major elements of the mantle rocks, aside from volatiles, are Mg (22.8 wt.%), Si (21.0 wt.%), Fe (6.26 wt.%), Ca (2.53 wt.%), and Al (2.35 wt.%) (McDonough and Sun, 1995). The composition of the mantle is also constrained by its density, seismic velocities, seismic anisotropy, and the compositions of ophiolite complexes and mantle xenoliths brought to the surface by kimberlitic and alkali basaltic eruptions. The elements of the mantle are contained in the minerals olivine ($(\text{Mg}, \text{Fe})_2\text{SiO}_4$), orthopyroxene ($(\text{Mg}, \text{Fe})\text{SiO}_3$), clinopyroxene ($[\text{Ca}, \text{Mg}]_2\text{NaAlSi}_2\text{O}_6$), and garnet

$((\text{Mg}, \text{Fe}, \text{Ca})_3\text{Al}_2\text{Si}_3\text{O}_{12})$. Olivine is a complete solid solution of Mg and Fe silicates with end members fayalite (Fe_2SiO_4) and forsterite (Mg_2SiO_4). Orthopyroxene is a limited solid solution with magnesium end member enstatite (the Fe end member is unstable). Clinopyroxene is a pyroxene solid solution with Ca and Al. Among all possible rock assemblages of these minerals only peridotites (olivine + pyroxene) and eclogites (pyroxene + garnet) are commonly found in mantle-derived samples. Eclogite is isochemical with basalt and transforms to basalt at depths of less than about 80 km in the Earth. The mineralogy and composition of the Earth's mantle might be that of pyrolite, a peridotite model introduced by Ringwood (1975) to explain the seismic, petrologic, and mineralogic properties of the upper mantle.

10.02.4.4 Mineralogy

Olivine and pyroxene transform to higher-density polymorphs under the high temperatures and pressures encountered at depth in the Earth's mantle. At a depth of about 410 km olivine transforms to spinel (the transformation occurs in two steps involving wadsleyite, or β -spinel, and at higher pressure ringwoodite, or γ -spinel). A prominent seismic discontinuity at the depth of about 410 km is believed associated with the olivine–spinel phase change. This subsolidus phase change is exothermic and involves jumps in seismic velocities and density. Another major seismic discontinuity occurs at the depth of about 660 km and this is believed to coincide with the transformation of ringwoodite to magnesium perovskite and magnesiowüstite. This solid–solid phase transformation is endothermic. The region between 410 km depth and 660 km depth is known as the transition zone. The depth of 660 km marks the boundary between the upper mantle and the lower mantle. Perovskite-forming reactions also occur in the pyroxene system but over a wider pressure interval than in the olivine system. Garnet, for example, dissociates to form MgSiO_3 and CaMgSiO_3 perovskites plus Al_2O_3 . The lower mantle is thus dominated by silicate perovskites.

Recent laboratory and theoretical results have demonstrated that perovskite transforms to still another structure, post-perovskite, at pressures and temperatures found just above the core–mantle boundary in the Earth (Murakami *et al.*, 2004; Oganov and Ono, 2004; Tsuchiya *et al.*, 2004). The perovskite–post-perovskite transformation, like the

olivine–spinel phase change, is an exothermic reaction. The location of this phase transition is coincident with the several hundred kilometer thick layer at the bottom of the mantle known as the D''-layer wherein seismic velocities undergo large variations (Lay *et al.*, 2005).

The subsolidus phase changes in the Earth's mantle have important effects on the dynamics of the mantle (Schubert *et al.*, 2001). The exothermic reactions generally promote convection while the endothermic phase change retards it. Do similar phase changes occur in other terrestrial planets? Venus is large enough that the olivine–spinel and spinel–perovskite phase changes should be present in its mantle. However, the pressure at the base of Venus's mantle may not be sufficiently high for the occurrence of the perovskite–post-perovskite phase change. Mars is so small that only the olivine–spinel phase change might occur near the base of its mantle. Possible phase changes in the mantles of the other planets will be discussed in more detail below.

10.02.5 The Moon

The Moon is one of the best-understood bodies in the solar system. The first billion years of the impact history of the Earth–Moon system are retained in the lunar crater record. The chronology of major impact events has been derived from rock and soil samples collected during the Apollo and Luna missions and represents the current-best database for dating other planetary surfaces in the solar system (Kaula *et al.*, 1986). A number of lunar meteorites provide additional constraints on the early evolution and bulk composition of the lunar crust (Korotev, 2005).

10.02.5.1 General

The Clementine and Lunar Prospector missions have returned global data sets of lunar gravity, topography, remanent magnetism, mineralogy, and chemical composition of the surface (Nozette *et al.*, 1994; Binder, 1998). The low mean lunar density implies that the Moon is depleted in iron relative to the other terrestrial planets. Dynamical modeling and isotopic data suggest that accretion and differentiation occurred soon after the formation of a vapor and debris cloud caused by the giant impact of a Mars-sized planetesimal into the early Earth. The internal differentiation of the Moon was accompanied by the formation of a magma ocean subsequent to hot

accretion and extraction of a highly aluminous flotation crust enriched in plagioclase feldspar (Wood *et al.*, 1970; Smith *et al.*, 1970; Warren, 1985). The duration and depth of differentiation during the magma-ocean phase, the possible existence of an undifferentiated lower mantle, and the mechanism of core formation are among the key questions of lunar science (Jolliff *et al.*, 2000b). For review articles on the internal constitution and thermal–magmatic evolution of the Moon, we refer the reader to Wieczorek *et al.* (2006) and Shearer *et al.* (2006), respectively.

10.02.5.2 Interior Structure

Models of the layering and composition of the lunar interior are widely based on the Apollo lunar seismic data combined with measurements of the mean density and the MoI factor. Seismic measurements at the Apollo 12, 14, 15, and 16 landing sites indicate that the lunar interior consists of a crust and a mantle. The chemical composition of the silicate mantle is consistent with that of an olivine–pyroxene mixture (Toksöz *et al.*, 1974; Hood and Jones, 1987). From Doppler tracking of the Lunar Prospector spacecraft and lunar laser ranging data on lunar libration, Konopliv *et al.* (1998) have inferred an improved average MoI factor of $I_p/M_p r_p^2 = 0.3931 \pm 0.0002$ and Konopliv *et al.* (2001) have derived a lunar tidal potential Love number of $k_2 = 0.026 \pm 0.003$. Despite the improved values of the moment of inertia and tidal potential Love number, models of the lunar interior are nonunique. Considerable uncertainty is connected with the radius and physical state of a possible metallic core, more gradual or discontinuous seismic velocity variations within the lunar mantle, and lateral and vertical heterogeneities of the lunar crust, the mean thickness of which is estimated at 49 ± 15 km if an Airy-type compensation mechanism applies (Wieczorek *et al.*, 2006).

Lunar laser ranging data show that the true spin axis of the Moon is displaced from the Cassini alignment (mean direction of the spin axis) by 0.26 arcsec. This can be explained by internal dissipation in the presence of a fluid core (Yoder, 1981; Dickey *et al.*, 1994; Williams *et al.* 2001). Furthermore, the paleomagnetic record of some lunar samples suggests the former existence of a significant magnetic field produced by dynamo action in a liquid metallic core (Hood and Jones, 1987). However, the size of a possible lunar core is less well determined. The seismic data allow the existence of a lunar core with

a radius of 170–360 km (Nakamura *et al.*, 1974). The MoI factor determined by Konopliv *et al.* (1998) is consistent with a core radius between 220 and 450 km. Furthermore, independent observations of the lunar magnetic moment induced in the geomagnetic tail of the Earth suggest that the lunar core radius is 340 ± 90 km (Hood *et al.*, 1999). A combined analysis of mean density, MoI factor, tidal potential Love number k_2 , and quality factor Q and the inversion of the entire set of lunar laser ranging and seismic data implies a molten or partially molten iron core with radius and density of about 350 km and 7200 kg m^{-3} , respectively (Khan *et al.*, 2004). The possible range of density models that simultaneously satisfy the mean density and moment of inertia of the Moon are shown in Figure 6.

10.02.5.3 Composition

The lunar surface is divided into light-colored heavily cratered highlands and smooth dark lowland maria which are most prominent on the near side. The highlands are saturated with large craters owing to their greater age in comparison to the maria and dominate the lunar far side and most of the near side. Highly anorthositic rocks are exposed in the lunar highlands, whereas the maria represent floods of basaltic lava that were erupted about 0.5 Gyr after major impact events. Little mare volcanism is associated with the South Pole Aitken basin, and Oceanus Procellarum, the largest expanse of mare volcanism,

is not necessarily associated with an impact event (Neumann *et al.*, 1996).

It appears that the majority of the lunar basalts erupted within the Procellarum-KREEP-Terrane (Jolliff *et al.*, 2000a), a unique geologic province that contains elevated abundances of heat-producing elements. Wieczorek and Phillips (2000) have shown that the enhanced heat production of this province could have melted the underlying mantle. Wieczorek *et al.* (2001) have suggested that mare basalts might preferentially erupt within the impact basins since mare basalts are more dense than the upper anorthositic crust, but less dense than the deep, more mafic, lower crust. If an impact event stripped away the upper crust, mare basalts could easily rise through the crust and erupt based solely on buoyancy considerations. The compositions of the mare basalts are consistent with volcanic source regions of several hundred kilometers depth (Heiken *et al.*, 1991). Mare basalt samples provide an assessment of the Mg number $\text{MgO}/(\text{MgO} + \text{FeO})$ of 0.75–0.8 and a bulk Al_2O_3 content of 1 wt.% for the upper mantle (Hood and Zuber, 2000). The ages of the basaltic maria, as determined from returned samples and careful crater chronologic studies, range between about 4 and ≈ 1.2 Gyr (Hiesinger *et al.*, 2003) and suggest that U-, Th-, and K-rich residua of the crystallized magma ocean providing the heat required for mantle melting were at least locally present (Jolliff *et al.*, 2000a; Wieczorek and Phillips, 2000). The lateral and depth variation of crustal composition has been assessed from compositional mapping of central peaks (Wieczorek and Zuber, 2001) and ejecta blankets of large impact basins (Bussey and Spudis, 2000). Deeply excavated ejecta are found to be more mafic than the surface material (Pieters and Tompkins, 1999).

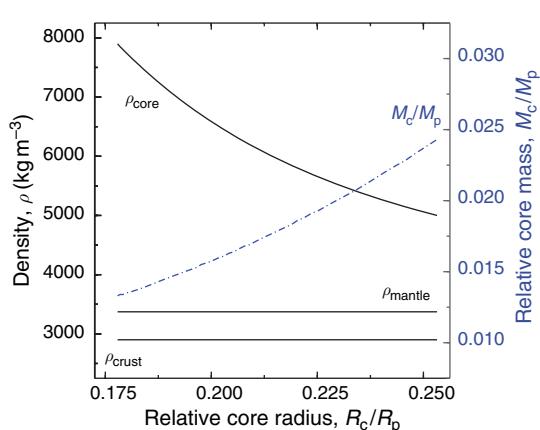


Figure 6 Three-layer model determinations of lunar mantle and core density and core mass fraction M_c/M_p vs relative core radius R_c/R_p consistent with the mean density and MoI factor of the Moon. The thickness and density of the anorthositic lunar crust is fixed at 60 km and 2900 kg m^{-3} , respectively.

10.02.5.4 Lunar Crust

The circular maria are frequently associated with mass concentrations (mascon basins) causing a positive anomaly in the lunar gravity field due to the larger density of the basalt layer compared to that of the surrounding anorthositic crust. Uplift of the crust–mantle interface provides another contribution to the positive gravity anomaly. Some mascon basins are known that do not appear to be associated with mare volcanism (Konopliv *et al.*, 1998, 2001). Neumann *et al.* (1996) and Wieczorek and Phillips (1999) have suggested that the lunar mascons might

partly result from superisostatic uplift of the crust–mantle interface.

The dynamical range of the lunar topography varies between 5 and 6 km on the near side and 16 km on the far side due to the presence of the 2500-km diameter and 8-km deep South Pole Aitken basin (Zuber *et al.*, 1994). This ancient basin was formed by a huge impact that might have penetrated through the crust into the upper mantle thereby creating a low-albedo mafic anomaly. However, using five-color images obtained by the Clementine spacecraft, Pieters *et al.* (1997) argue that the mineralogy of the mafic component represents lower crust rather than upper-mantle material. Wieczorek and Phillips (1998) also suggest that there is still a 10–20-km-thick layer of crustal material present within the basin.

There is a pronounced dichotomy between the near-side and far-side crustal thickness as a consequence of the early differentiation of the Moon. Several factors such as variable heat transfer in the molten lunar interior and the large-scale insulation due to protocrust formation and thick impact ejecta may have contributed to the asymmetry (Jolliff *et al.*, 2000b). An early analysis of body-wave phases from artificial impacts of known impact time and location provided evidence for a dual-layered crust about 60 km thick beneath the Apollo 12 and 14 stations near Mare Cognitum (Toksoz *et al.*, 1974), whereas a putative thickness of 75 ± 5 km was derived from seismic data near the Apollo 16 highland site (Goins *et al.*, 1981). Previous estimates of the crustal thicknesses varied from 30–35 km under mascon basins to 90–110 km beneath the highlands, whereas irregular maria have intermediate thickness values of 50–60 km (Bills and Ferrari, 1977). From a re-analysis of gravity data acquired by the Apollo and Clementine spacecraft and assuming a thickness of 55 km at the Apollo 12/14 site, the thickness of the lunar farside crust has been estimated at 67 km. This would result in a mean thickness of 61 km if a uniform crustal composition is assumed (Neumann *et al.*, 1996).

A difference in crust thickness between near side and far side is thought to contribute substantially to the 1.7 km offset of the CoM relative to the center of figure CoF of the Moon in the direction of Earth (Kaula *et al.*, 1974; Hood and Jones, 1987; Zuber *et al.*, 1994). Wieczorek and Phillips (1998) have computed a variety of crustal thickness maps for the Moon, assuming both homogeneous and dual-layered crusts. The homogeneous crust model

characterized by a constant crustal density provides a total crust thickness of 66 km, whereas the preferred dual-layered model of the lunar crust consisting of a 31-km-thick upper crust and a 29-km-thick lower crust yields a total thickness of 60 km. The gravitational field measurements of the Lunar Prospector spacecraft suggest a mean crustal thickness of about 70 km assuming Airy compensation of the lunar highlands (Konopliv *et al.*, 1998) (*see* Chapter 10.05).

10.02.5.5 Lunar Seismology and Mineralogy

The Moon is the only planetary body other than the Earth for which a seismic velocity structure has been derived from analyses of the Apollo seismic data set (*see* Chapter 10.03). A seismic network of four stations on the lunar near side, installed at the landing sites of Apollo 12, 14, 15, and 16 from 1969 to 1972, continued to operate until it was turned off in September 1977. The seismic activity of the Moon is provided by deep moonquakes located at 850–1000 km depth, shallow moonquakes or high-frequency teleseismic events situated at 50–220 km depth, and meteoroid impacts onto the lunar surface. The deep moonquakes, apparently confined to the lunar near side, are most numerous and triggered by tidal deformation of the Moon. More than 12 000 events associated with 81 identified sources ('nests') have been recorded during the lifetime of the Apollo seismic network. It is likely that about 30 nests are located on the lunar far side. This implies that either the Moon's deep interior within about 40° from the sub-Earth antipode is nearly aseismic or, alternatively, that the region of the deep lunar interior deflects seismic waves and is highly attenuating for S waves, thereby impeding the localization of far side moonquake source regions (Nakamura, 2005).

Early seismic analyses were based on a limited set of arrival-time readings (Goins *et al.*, 1981), whereas the model of Nakamura (1983) employs the complete data set of 5-year simultaneous operation of four Apollo seismometers including more deep moonquake sources. More detailed lunar velocity and density structure models have been obtained from a new inversion of different subsets of the Apollo seismic data set by applying various inversion techniques to revised arrival-time readings (Khan *et al.*, 2000; Khan and Mosegaard, 2002; Khan *et al.*, 2004; Lognonné *et al.*, 2003; Chenet *et al.*, 2006).

The multilayer model of Goins *et al.* (1981) consists of a dual-layer crust, an upper mantle, an intermediate transition zone, and a lower mantle.

The crust is subdivided into a 20-km-thick upper layer with P-wave velocity $V_P = 5.1 \text{ km s}^{-1}$ and S-wave velocity $V_S = 2.96 \text{ km s}^{-1}$ and underlain by a lower-crust layer with $V_P = 6.8 \text{ km s}^{-1}$ and $V_S = 3.9 \text{ km s}^{-1}$ varying between 20 and 60 km in thickness. The upper mantle then extends from 60 to 400 km depth with average seismic velocities of $V_P = 7.7 \pm 0.15 \text{ km s}^{-1}$ and $V_S = 4.45 \pm 0.05 \text{ km s}^{-1}$. The moderate velocity decrease with depth is believed to be related to the temperature increase with depth. The corresponding seismic quality factors are $Q_P = 5000$ and $Q_S = 3000$, respectively. A narrow transition zone is situated between 400 and 480 km depth and reveals a prominent decrease of the shear-wave velocity, thereby creating a seismic shadow zone. The substantial decrease of the shear-wave velocity at the base of the transition zone has been interpreted in terms of a compositional change to a more primitive undifferentiated interior. The lower mantle ranges from 480 to 1100 km depth with average seismic velocities of $V_P = 7.6 \pm 0.6 \text{ km s}^{-1}$ and $V_S = 4.2 \pm 0.1 \text{ km s}^{-1}$ and corresponding seismic quality factors of $Q_P = 1500$ and $Q_S = 1000$, respectively. However, significant shear-wave attenuation accompanied by a Q_S drop by one order of magnitude occurs below a depth of 1100 km and suggests that the lowermost lunar mantle is at least partially molten.

The model of Nakamura (1983) includes a four-layer crustal structure that was introduced to satisfy the P-wave traveltimes of artificial impacts. The lunar mantle is subdivided into three layers with discontinuities at 270 and 500 km depth. The average velocities in the upper mantle above 270 km depth decrease from $V_P = 7.74 \text{ km s}^{-1}$ and $V_S = 4.49 \text{ km s}^{-1}$ to $V_P = 7.46 \text{ km s}^{-1}$ and $V_S = 4.25 \text{ km s}^{-1}$ between 270 and 500 km depth. In contrast to earlier studies based on a limited data set, the average velocities in the middle mantle between depths of 500 and 1000 km are about 8.26 and 4.65 km s^{-1} , respectively. The relatively high velocities in the middle mantle layer may represent Mg-rich olivine rock left over from early melting and differentiation of the deep lunar interior. The multilayer lunar interior model suggested by Nakamura (1983) is an approximation to more realistic continuous models with P and S velocities gradually decreasing with depth.

Contrary to previous studies, mean crustal thickness of only $45 \pm 5 \text{ km}$ (Khan *et al.*, 2000) or $38 \pm 3 \text{ km}$ (Khan and Mosegaard, 2002) have been proposed based on a re-analysis of Apollo lunar seismic data (Figure 7). The P-wave velocity increases from the

surface to the crust–mantle transition. The upper mantle extends from 45 to 500 km depth with average P- and S-wave velocities of $8.0 \pm 0.8 \text{ km s}^{-1}$ and $4.0 \pm 0.4 \text{ km s}^{-1}$, respectively. In contrast to the model of Nakamura (1983), the upper mantle is characterized by constant seismic velocities indicating a homogeneous structure and composition. Similar to the model of Nakamura (1983), the transition to the middle mantle at a depth of $560 \pm 15 \text{ km}$ is marked by a sharp increase to $V_P = 9.9 \pm 1.9 \text{ km s}^{-1}$ and $V_S = 5.9 \pm 0.9 \text{ km s}^{-1}$. It should be noted, however, that such high seismic velocities are difficult to reconcile with any known mineral phase assemblages. An intermediate low-velocity layer may cover the depth range from 620 to 700 km. Below that depth, extends a high-velocity layer with P- and S-wave velocities of 11.0 ± 2.1 and $6.0 \pm 0.7 \text{ km s}^{-1}$, respectively, encompassing the deep moonquake source region.

Using arrival-time readings of about 60 deep and shallow lunar quakes and meteoroid impacts, Lognonné *et al.* (2003) propose crust thicknesses in the range $30 \pm 2.5 \text{ km}$ at the Apollo 12/14 site situated in the Procellarum-KREEP-Terrain. The corresponding temperature range of the underlying pyroxenite-rich mantle compares well to electrical conductivity profiles retrieved from early Apollo orbiter observations (Khan *et al.*, 2006b). The mantle temperatures are still sufficient to keep an Fe–FeS core molten but not a pure iron core. Chenet *et al.* (2006) limit themselves to a seismic analysis of seven artificial and 19 natural usable meteoroid impacts suggesting only moderate lateral variations in lunar near-side crust thickness. The crust thickness beneath the Apollo 12, 14, 15, and 16 landing sites is found to be 33 ± 5 , 31 ± 7 , 35 ± 8 , and $38 \pm 7 \text{ km}$, respectively, consistent with a combined gravity/topography analysis. Considering a mean crust density and an upper mantle P-wave velocity of $\rho = 2900 \text{ kg m}^{-3}$ and $V_P = 7.57 \text{ km s}^{-1}$, respectively, the lunar anorthositic crust would be much thinner than previously thought with a mean thickness of only $40 \pm 5 \text{ km}$ (Chenet *et al.*, 2006). Assuming an anorthositic composition for the 40-km-thick crust, Lognonné *et al.* (2003) find the Moon's silicate portion (mantle + crust) to be composed of 53.5 wt.% SiO_2 , 6.4 wt.% Al_2O_3 , 13.3 wt.% FeO , 21.9 wt.% MgO , and 4.9 wt.% CaO with a bulk uranium content of 28 ppb.

The most prominent increase in seismic velocity that occurs at a depth of $\sim 500 \text{ km}$ represents the transition from the upper mantle to the middle mantle. This discontinuity has been interpreted in terms of a mineralogic phase transition from the spinel to

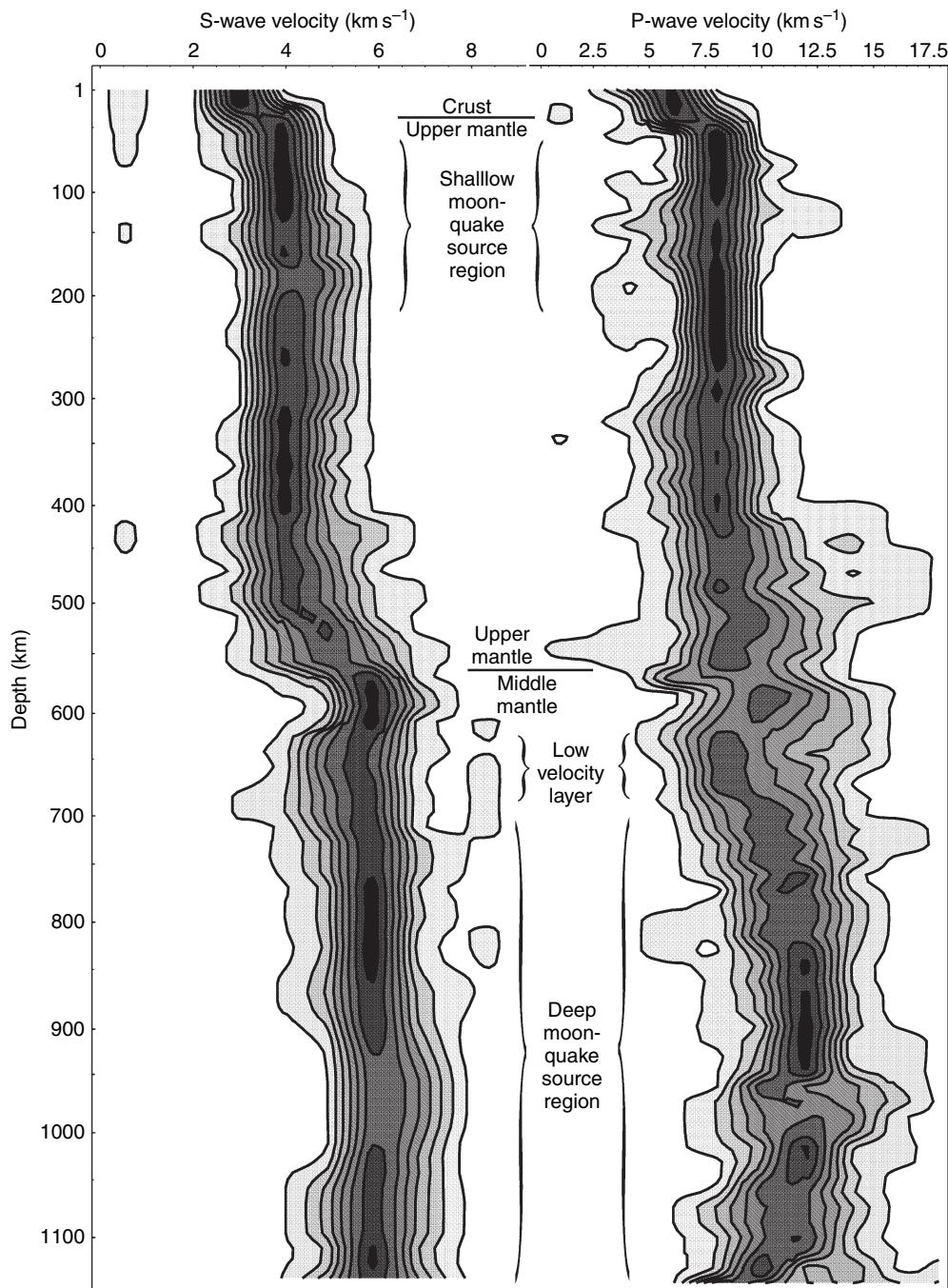


Figure 7 Marginal posterior probability distributions illustrating the range of possible (left) S-wave and (right) P-wave velocity structures of the Moon based on a total number of 50 000 models. The contour lines define nine equally sized probability density intervals for the distributions. Note that maximum probability does not necessarily correspond to maximum likelihood of a calculated velocity structure. Adapted from Khan A and Mosegaard K (2002) An inquiry into the lunar interior: A nonlinear inversion of the Apollo lunar seismic data. *Journal of Geophysical Research* 107(E6): 50036 (doi:10.1029/2001JE001658).

the garnet stability field. Alternatively, a change in composition to more aluminous and MgO-rich mafic silicates has been invoked thereby increasing the Mg

number below this depth (Nakamura, 1983; Khan *et al.*, 2006a). The re-analysis of the Apollo lunar seismic data indicates a homogeneous, constant-

velocity upper mantle extending down to 560 ± 15 km depth, whereas the radial velocity distribution suggests more inhomogeneous middle-and lower-mantle layering (Khan *et al.*, 2000; Khan and Mosegaard, 2002). The compositional change across the discontinuity together with the homogeneity of the upper mantle and the inhomogeneity of the middle mantle has been interpreted in terms of the initial depth of melting and differentiation during the magma-ocean phase of the Moon (Hood and Zuber, 2000). Based on mare-basalt petrology and thermal evolution considerations, Elkins-Tanton *et al.* (2003a) identify the 500-km discontinuity with the maximum depth of melting beneath the Procellarum-KREEP-terrane on the lunar near side. It is also possible that the 500-km discontinuity represents stratified olivine- and orthopyroxene-rich cumulates that were subsequently emplaced at the bottom of the lunar magma ocean. The mineralogical layering is supported by seismic inversions that account for thermodynamic mineral phase equilibria (Kuskov and Fabrichnaya, 1995; Kuskov, 1995, 1997; Kuskov and Kronrod, 1998).

The existence of a small lunar core could not be confirmed or rejected solely on the basis of the seismic data acquired by the Apollo missions. During the limited operational period of the lunar seismic network with stations on the near-Earth hemisphere, only one meteoroid event was detected with seismic rays crossing a central low-velocity zone from which the size of a lunar core was estimated for the first time. However, Sellers (1992) has shown that a relocation of this meteorite impact on day 263 in 1973 can explain the arrival-time and signal-amplitude data without requiring the presence of a lunar core. From tentative arrival-time determinations of two of the largest, most distant impacts it has been found that the corresponding seismic phases could have traveled at a P-wave velocity of about 5 km s^{-1} through a lunar core about 400–450 km in radius (Sellers, 1992). Considering the ambiguity inherent in the Apollo lunar seismic data, first reliable recordings of seismic phases traveling through the central region of the Moon are therefore expected from future measurements, like those once envisaged for the Japanese Lunar-A mission (Mizutani 1995).

10.02.6 Mercury

The planet Mercury is the least well-known body among the terrestrial planets and is unique in many respects. It represents an end member of the

terrestrial planets with respect to its density and distance from the Sun and thereby provides important constraints on planetary formation and evolution in the innermost part of the solar nebula (Balogh and Giampieri, 2002; Solomon, 2003).

10.02.6.1 General

Mercury is 4878 km in diameter, roughly one-third the size of Earth, and occupies only about 6% of the volume of Earth. Albeit substantially smaller in size, Mercury's surface gravity of 3.7 m s^{-2} is like that of the larger planet Mars. In 1974 and 1975, the Mariner 10 spacecraft provided the first close-up look of the planet during three encounters. From these flybys, better values of the planet's mass, mean radius, and average density were obtained. Less than half of the surface was covered by images at an average resolution of 1 km, and less than 1% at 100–500 m resolution, resulting in a limited characterization of surface morphology and geological evolution. Among the most important findings of the Mariner 10 spacecraft was the then unexpected detection of a magnetic field of internal origin (Ness, 1979).

The densely cratered surface of Mercury suggests that it is one of the oldest surfaces in the solar system, but not necessarily older than the lunar and Martian highlands. About 60% of the known surface is covered by intercrater and smooth plains possibly caused by volcanic emplacement. Based on the notion that the Mercurian plains closely resemble the lunar light plains, they might also be formed of impact materials, perhaps impact melt or other basin ejecta that behaved more like a fluid when emplaced (Wilhelms, 1976). Lobate scarps and widespread volcanic plains suggest an early evolution in which volcanically induced expansion preceded a phase of planetary contraction (Solomon, 2003).

Mercury's orbit about the Sun has a semimajor axis of 0.387 AU, an eccentricity of 0.206, and an inclination of 7° relative to the ecliptic. Therefore, the orbital distance from the Sun varies between 0.308 AU at perihelion and 0.466 AU at aphelion in the course of one revolution, causing significant tidal distortion of the planet (Burns, 1976; Van Hoolst and Jacobs, 2003). Furthermore, the near-surface layer of Mercury is exposed to a severe thermal environment in terms of elevated surface temperatures and high subsurface temperature gradients due to the planet's proximity to the Sun with an insolation of up to 15 kW m^{-2} . Related diurnal surface temperature variations between 90 and 740 K are greater than on

any other planet or satellite in the solar system (Strom, 1997). Radar observations from the Earth have shown that the rotation period (58.6 d) is locked into a 3:2 resonance with the orbital period (87.7 d). As a consequence of the 3:2 spin–orbit coupling caused by tidal interactions with the Sun, one solar day on Mercury lasts 176 Earth days and corresponds to two revolutions or three rotations of the planet. While the distribution of solar irradiation is symmetric between both hemispheres, the significant orbital eccentricity causes longitudinal variations superimposed on the latitudinal variation of the solar irradiation (Van Hemelrijck and Vercheval, 1981). Due to the lack of a substantial atmosphere, the surface of Mercury has been heavily cratered and fragmented by impacts and small-particle bombardment resulting in a planet-wide, fine-grained, and thermally insulating regolith layer (Langevin, 1997). The thermal and electrical properties of the Mercurian regolith derived from Mariner 10 measurements are similar to those of the Moon (Chase *et al.*, 1976). Even large-amplitude surface temperature perturbations are thus expected to rapidly fade away with depth and will be negligible below several thermal skin depths (Vasavada *et al.*, 1999; Hale and Hapke, 2002).

10.02.6.2 Interior Structure

The large average density of $5430 \pm 10 \text{ kg m}^{-3}$ (Anderson *et al.*, 1987) is comparable to that of the Earth and Venus but much larger than that of the Moon and Mars. The corresponding zero-pressure density of about 5300 kg m^{-3} is even much higher than the uncompressed densities of Earth, Venus, and Mars which are about 4100, 4000, and 3800 kg m^{-3} , respectively. This suggests that Mercury contains a larger proportion of heavier elements such as iron than any other terrestrial planet. The mass concentration of iron should be about twice that in the Earth (Wasson, 1988). The existence of the weak intrinsic magnetic field and compressional surface features observed by Mariner 10 together with the large average density suggest that most of the iron is concentrated in a substantial Fe-rich core with a radius of roughly 0.8 times the planet radius (Figure 8). The core occupies about 42% of the planet's volume corresponding to a core mass fraction of two-thirds relative to the planet's mass or about twice that of the Earth (Siegfried and Solomon, 1974). If Mercury formed by the accretion of high-temperature condensates close to the Sun then a substantial depletion of volatiles such as sulfur should have occurred. This would translate into a refractory

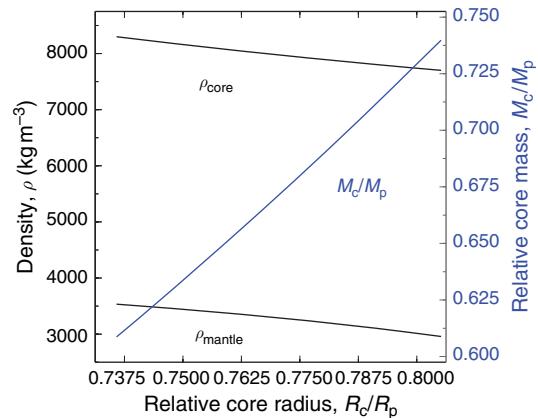


Figure 8 Two-layer model determinations of mantle and core density and core mass fraction M_c/M_p of Mercury vs relative core radius R_c/R_p based on a Mol factor of 0.3359. The core size of Mercury could be relatively well determined if the planet's moment of inertia becomes known.

bulk composition and a metallic core predominantly composed of iron with only small amounts of sulfur (Lewis, 1988; Goettel, 1988).

Radar ranging data suggest that the equatorial shape of Mercury is highly elliptical. However, since Mercury has not attained an equilibrium figure, its shape and gravitational field cannot be used to infer the size of its metallic core. There is a significant equatorial ellipticity $(a - b)/a = (540 \pm 54) \times 10^{-6}$. The CoF is offset from the CoM by $640 \pm 78 \text{ m}$ in the direction of the still unseen hemisphere of Mercury (Anderson *et al.*, 1996). The CoF–CoM offsets of all the terrestrial planets are plausibly attributed to hemispheric asymmetries in crustal thickness depending on the contrast $(\rho_m - \rho_s)/\rho_m$ between the densities of the mantle and the crust ρ_m and ρ_s , respectively. The magnitude of the CoF–CoM offset implies an excess crustal thickness of $< 12 \text{ km}$, which is comparable to that obtained for the Moon. From a comparison between the equatorial shape and the gravitational equatorial ellipticity $C_{2,2}$ as inferred from the Mariner 10 flybys, Anderson *et al.* (1996) have concluded that the Mercurian crust could be $200 \pm 100 \text{ km}$ thick if Mercury's equatorial ellipticity were entirely compensated by Airy isostasy. Even the lower bound of 100 km is more than twice as large as the crustal thickness of the other terrestrial planets and the Moon and, therefore, difficult to reconcile with the planet's magmatic history if the Mercurian crust formed from volcanic emplacement of partial melts extracted from the underlying mantle.

10.02.6.3 Composition

Models of the interior structure rely on the mass and mean radius of the planet since a value for the moment-of-inertia factor is not available at present. A determination of the MoI factor, as envisioned by future missions to Mercury, would help distinguish an iron core from a more homogeneous distribution of iron in oxidized form within the planet (Schubert *et al.*, 1988). The spectral characteristics and high albedo of the surface of Mercury are consistent with the existence of a metal-poor and possibly highly differentiated, feldspathic crust that contains less FeO and TiO₂ ($\sim 2\text{--}4\%$) than the lunar highland crust ($>15\%$) (Sprague *et al.*, 1994; Blewett *et al.*, 1997; Warell and Blewett 2004). This is taken as evidence for the strong internal differentiation of the planet and implies the former existence of a magma ocean (Jeanloz *et al.*, 1995). Additional assumptions about the chemistry and densities of a basaltic crust, a more primitive mantle, and an iron-rich core are then required to construct models of the interior in accordance with the mass and mean density of the planet. Possible interior structure models have been calculated for various bulk chemical compositions based on different condensation and accretion scenarios (e.g., Wood *et al.*, 1981). Accounting for pressure-induced compression and thermal expansion effects, Siegfried and Solomon (1974) and Harder and Schubert (2001) propose MoI factors in the range from 0.325 for fully differentiated models to 0.394 for chemically homogeneous, undifferentiated models. The silicate shell comprising crust and mantle layers is then 500–700 km thick. There is little variation of pressure, temperature, density, and elastic moduli with depth through the silicate shell (Figure 9). Density discontinuities induced by major phase transitions should not be present in the mantle due to small pressure increase with depth resulting in a pressure at the core–mantle boundary of about 7 GPa (Siegfried and Solomon, 1974; Harder and Schubert, 2001). If a crust layer were added with thickness and density of 50 km and 3000 kg m^{-3} , respectively, the radius of a large iron core would be about 1800–1900 km thereby corresponding to a MoI factor of about 0.34 (Figure 10). It is possible, however, that compositional changes occur across the mantle. Thermal history calculations show that an asthenosphere in the upper mantle may be sustained up to the present time (Conzelmann, 1994). This has important consequences for the magmatic history of the planet.

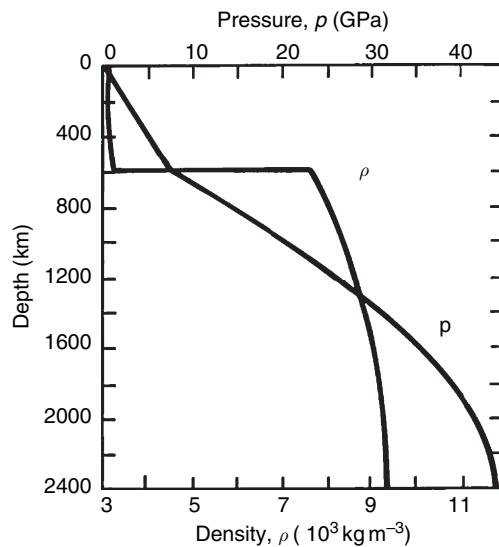


Figure 9 Variation of pressure p and density ρ versus depth for a fully differentiated model of Mercury's interior. Note that a refractory bulk composition is assumed (Schubert *et al.*, 1988). Adapted from Siegfried RW and Solomon SC (1974) Mercury: Internal structure and thermal evolution. *Icarus* 23: 192–205.

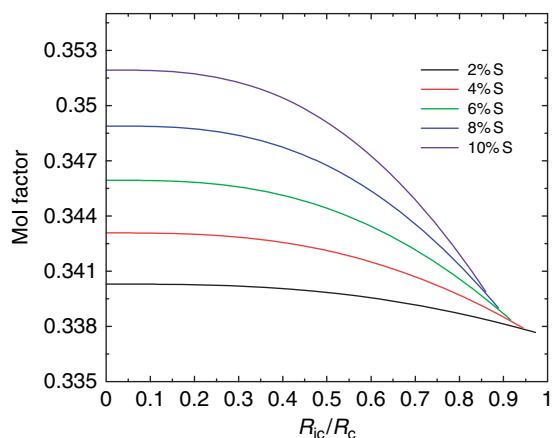


Figure 10 Mean Mol factor of a three-layer model of Mercury as a function of inner-core radius R_{ic} and core chemistry. R_{ic} is given relative to the radius of the core R_c .

10.02.6.4 Magnetic Field

Mercury is the only terrestrial planet other than the Earth with a perceptible dipole magnetic field. The presence of an internally generated magnetic field suggests that the iron core is at least partially liquid with an electrically conducting outer core of unknown thickness surrounding a solid inner core. Thermal evolution models indicate that the core

would have solidified early in the history of Mercury unless a light alloying element such as sulfur were present. A small amount of sulfur as suggested by Stevenson *et al.* (1983) is sufficient to depress the freezing point of a core alloy and is consistent with the refractory bulk composition as derived from the planet's condensation and accretion in the hot innermost part of the solar nebula. The freezing of an inner pure iron core would be accompanied by the enrichment of sulfur in the outer core thereby increasing the depression of the freezing point and maintaining a liquid outer-core shell in spite of planet cooling (**Figure 11**). Thermal history models taking into account parametrized convective heat transport through the mantle indicate sulfur concentrations of 1–5 wt.% to retain a liquid outer core shell at the present time (Stevenson *et al.*, 1983; Schubert *et al.*, 1988; Spohn, 1991). Based on parametrized models of the coupled thermal, magmatic, and tectonic evolution of Mercury, Hauck *et al.* (2004) conclude that a dry-olivine mantle rheology, thorium-dominated radiogenic heating suggesting late silicate-mantle vaporization (Cameron *et al.*, 1988), and a bulk core sulfur content of at least 6.5 wt.% is needed to explain both the planet's radial contraction of 1–2 km and the presence of a dynamo-driven, intrinsic magnetic field. Models of mantle convection including pressure and temperature-dependent rheology demonstrate that the cooling history of a terrestrial planet is governed by the growth of its lithosphere while the deep interior remains relatively hot. These models compare well to the parametrized convection

calculations but produce thicker outer core shells at identical sulfur concentrations. Depending on the stiffness of the mantle rheology, a liquid outer-core layer is then sustained even for sulfur concentrations as small as 0.2 wt.% consistent with cosmochemical arguments in favor of a volatile-poor planet (Conzelmann, 1994; Spohn *et al.*, 2001b). The 2-km radial contraction of Mercury in the absence of large-scale magmatism about 4 Gyr ago may be linked to core shrinkage due to solid-inner-core growth and mantle cooling governed by lithospheric thickening and sluggish mantle convection (Schubert *et al.* 1988).

10.02.6.5 Future Exploration

Future spacecraft missions together with Earth-based radar observations are expected to provide important new constraints on the internal structure of Mercury by determining its gravity field, large-scale topography, and tidal and rotational parameters with unprecedented accuracy. The ratio C_m/C between the axial moment of inertia of the planet's solid portion and that of the entire planet depends on the coupling between core and mantle and provides clues to the physical state of the outermost part of Mercury's core. C_m/C will be around 0.5 for a liquid core or liquid outer-core shell (**Figure 12**) and about 1 in case of a solid core. C_m/C can be obtained from accurate measurements of the planet's 88-day forced libration amplitude in longitude, the obliquity, and the second-degree coefficients C_{20} and C_{22} of the spherical harmonic expansion of the planet's

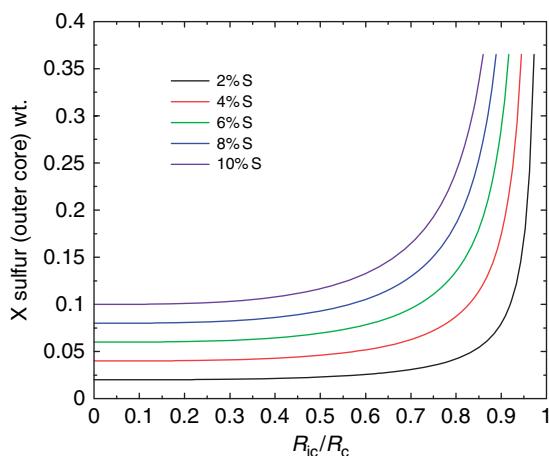


Figure 11 Sulfur content of the outer core of a three-layer model of Mercury as a function of inner core radius R_{ic} and core chemistry. R_{ic} is given relative to the radius of the core R_c .

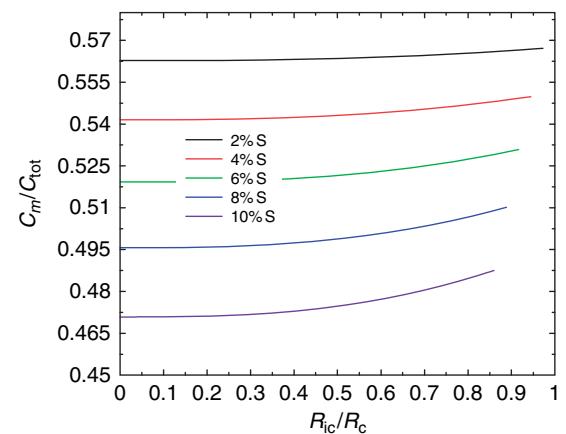


Figure 12 Ratio between the axial moment of inertia of Mercury's solid portion and that of the entire planet as a function of inner core radius R_{ic} and core chemistry. R_{ic} is given relative to the radius of the core R_c .

gravitational field (Peale, 1976, 1988) (see Chapter 10.04). C and C_m/C will permit calculation of the inner-core radius r_i and the outer core density for prescribed silicate mantle and solid-inner-core densities. Since Mercury is tidally flexed in its highly eccentric orbit about the Sun, the tidal Love number k_2 can be derived from time-variable gravitational field measurements on a spacecraft orbiting the planet (Van Hoolst and Jacobs, 2003). k_2 may provide useful constraints on the radial rigidity distribution and the extent of core differentiation for significant inner-core sizes, that is, $R_i/R_c > 0.5$ (Figure 13). Furthermore, the higher-order components of the gravitational field can be used to estimate crust thickness variations at short wavelengths and core–mantle boundary undulations at long wavelengths; the latter should be easily detectable due to the large size of Mercury’s core relative to the planet’s size (Spohn et al., 2001b).

No seismic data for Mercury exist. However, the periodic deformation of Mercury by solar tides may have important consequences for the planet’s seismic environment. Due to its highly elliptic orbit about the sun and its bound rotation, Mercury is exposed to strong tidal forces (Van Hoolst and Jacobs, 2003). As a result, the seismic environment of Mercury may feature large numbers of lunar-like quakes.

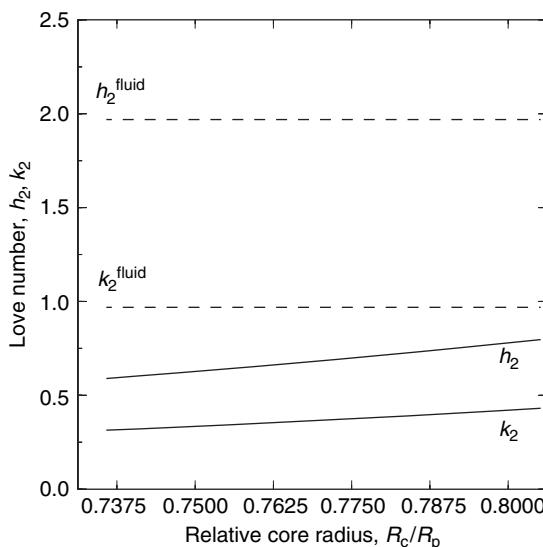


Figure 13 Tidal potential Love number k_2 of Mercury as a function of core radius R_c relative to the planet’s radius R_p . A two-layer structural model with a solid mantle and a fluid core is assumed. The fluid Love numbers k_2^{fluid} and h_2^{fluid} are related to the Mol factor fixed at a value of 0.3359 in these models.

10.02.7 Mars

The size of Mars is about half the size of the Earth and its mass is about one-tenth the mass of the Earth. The uncompressed density of roughly 3800 kg m^{-3} is significantly lower than the uncompressed densities of the Earth and Venus. The environmental conditions are the most earthlike among the terrestrial planets with surface temperatures varying between 140 K at night in winter and 300 K at midday in summer. Mars is thought to be a one-plate planet, as are the Moon and Mercury, lacking plate tectonics at least at the present time.

10.02.7.1 General

The Martian surface is characterized by the hemispheric dichotomy between relatively sparsely cratered lowland plains in the north and the heavily cratered southern highland terrain once subjected to the post-accretionary heavy bombardment of the inner solar system. The Tharsis rise is situated close to the near-equatorial boundary between the Northern and Southern Hemispheres and represents a giant volcanic dome established early in the planet’s history. Major volcanoes such as Olympus Mons and the Tharsis Montes were emplaced on top of the Tharsis rise. The magmatic evolution of the planet is characterized by a progressive concentration of the volcanic activity to the Tharsis area and, to a lesser extent, to the Elysium area located about 100° to the west in the northern lowland plains. The morphology of the boundary between the two hemispheres is dominated by outflow channels and chaotic terrain extending along broad gradual slopes rather than several-kilometer-high escarpments. Some portions of the dichotomy boundary are composed of fretted terrane and others are sculptured by scarps (Smith and Zuber, 1996). For review articles on the geophysics and evolution of Mars, we refer the reader to Schubert et al. (1992), Spohn et al. (1998, 2001a), and Solomon et al. (2005).

10.02.7.2 Interior Structure

The global topography and gravitational field of Mars have been determined with high accuracy using laser altimetry and two-way Doppler tracking of the Mars Global Surveyor (MGS) spacecraft (Smith et al., 1999b,a; Zuber et al., 2000; Lemoine et al., 2001). Further improvements of the gravity

model of Mars have been achieved by combining MGS tracking data with Mars Odyssey tracking data and surface tracking data from the Pathfinder and Viking 1 landers; the latter permits improvement of the parameters describing the orientation of the planet's rotational axis (Konopliv *et al.*, 2006). The rotational flattening of Mars results in a difference between the polar (north–south average) and equatorial radii of about 20 km (Seidelmann *et al.*, 2002). The variation of topography with respect to the Martian geoid is about 30 km, representing the largest dynamic range of any terrestrial planet. The hemispheric dichotomy of Mars is related to an offset of about 3 km between the planet's CoM and CoF along the polar axis. The Tharsis Bulge causes an additional offset of about 1.4 km along an equatorial axis in the direction of Tharsis. The topographic data show that the Tharsis rise consists of two broad rises, a larger, nearly circular southern rise superposed on the highlands that contains the Tharsis volcanoes and a smaller northern rise superposed on the lowlands that contains the shield volcano Alba Patera (Smith *et al.*, 1999a).

Whereas the gravitational field of the elevated Southern Hemisphere is relatively featureless and implies a state of near-isostatic compensation, the northern lowland plains reveal a wider range of gravitational anomalies (Smith *et al.*, 1999b). The dichotomy boundary is not clearly resolved on the gravity map, whereas the Tharsis Montes, Olympus Mons, Valles Marineris, and Isidis impact basin are visible as individual gravitational anomalies in the aeroid. Major gravity highs are associated with the Tharsis and Elysium volcanoes indicating that they are not isostatically compensated. Large impact basins reveal negative annular anomalies with a central positive anomaly. The global crust and upper mantle structure of Mars has been derived from *MGS* measurements of gravity and topography. The Bouguer gravity has been interpreted in terms of crustal thickness variations. In these models the thickness of the Southern Hemisphere crust decreases progressively from south to north, whereas the northern lowlands are characterized by a more uniform crust thickness (Zuber *et al.*, 2000; Neumann *et al.*, 2004).

Because of the nonuniqueness inherent in the interpretation of gravitational field data, unless they can be tied to direct seismic measurements, the mean density and thickness of the Martian crust are unknown at present. Plausible Martian crust densities range from 2700 to 3100 kg m⁻³ (Wieczorek and

Zuber, 2004) based on end-member-type compositional models of the Martian crust. The lower estimate of about 2700 kg m⁻³ represents an andesitic–basaltic composition obtained from Pathfinder-APXS measurements of soil-free Martian rocks (Brückner *et al.*, 2003). The upper estimate of about 3100 kg m⁻³ represents porous basaltic shergottites (Britt and Consolmagno, 2003) believed to be samples of the Martian crust released during one or several giant impacts (McSween, 1994). These rocks are further believed to represent end members of the composition of well-mixed Martian soil on a planetary scale (Nimmo and Tanaka, 2005). Using a spectral localization method to analyze volcanic surface loads, Belleguic *et al.* (2005) have reported considerably higher crustal densities beneath the Elysium rise of 3270 ± 150 kg m⁻³, whereas the load densities related to the major Martian volcanoes except Alba Patera are best constrained by a value of 3200 ± 100 kg m⁻³. The latter density is similar to that of Shergotty-type basaltic meteorites thought to have originated from Mars. This implies that the Martian lowland crust is composed of more mafic constituents than the southern highland crust and that the hemispheric difference in elevation may be mainly attributed to a Pratt-like isostatic compensation mechanism (Belleguic *et al.*, 2005). Present estimates of the mean crustal thickness of Mars are entirely based on indirect geophysical studies like local relationships between gravity and topography and/or geochemical arguments and may range between 30 and 80 km (Neumann *et al.*, 2004; Wieczorek and Zuber, 2004; Solomon *et al.*, 2005) although crust thicknesses of up to about 100 km are consistent with global geophysical constraints (Sohl and Spohn, 1997; Kavner *et al.*, 2001; Gudkova and Zharkov, 2004; Sohl *et al.*, 2005). Assuming a Bouguer reduction density of 2900 kg m⁻³ and a crust–mantle density contrast of 600 kg m⁻³, crust thicknesses vary from about 6 to 102 km, the former of which is located beneath the center of Isidis and the northwestern floor of Hellas Planitia (Neumann *et al.*, 2004). Larger crust thickness might be unlikely, however, since ductile flow in the warm lower crust could cause relaxation of lateral crustal thickness variations (Zuber *et al.*, 2000; Zuber, 2001; Nimmo and Stevenson, 2001). The crust thickness could be further limited by the pressure-induced transition from basalt to eclogite that may occur in the lower Martian crust (Babeyko and Zharkov, 2000).

Vector magnetic field observations, obtained at altitudes between about 100 and 200 km subsequent

to orbit insertion of the Mars Global Surveyor spacecraft, have been compiled into a global magnetic field map of the Martian crust (Connerney *et al.* 2005). Although Mars lacks an intrinsic magnetic field at present, the Martian crust acquired intense magnetization in the past, hinting at the existence of a strong internally generated field at that time. Crustal remanent magnetization has been found to exceed that of Earth by more than an order of magnitude and is mainly confined to the most ancient, heavily cratered southern highland terrain (Acuña *et al.*, 1999; Connerney *et al.*, 1999). Linear magnetic features of up to 2000 km in length are frequently oriented in the east-west direction, some resembling a pattern of band-like features of alternating magnetic polarity (Connerney *et al.*, 2001). The Martian crust may have been magnetized during a time span of only a few hundred million years after planet formation when an active core dynamo driven by thermal convection produced an intense global magnetic field. If crust magnetization occurred later in the planet's evolution, a core dynamo driven by chemical convection associated with inner-core growth would be more likely (Connerney *et al.*, 2004). The lack of crustal magnetization in the vicinity of large impact basins (Hellas, Argyre, Isidis) is commonly attributed to cessation of the core dynamo when these basins formed about 4 Gyr ago during the Noachian period and demagnetized underlying crust layers (Acuña *et al.*, 1999). Alternatively, however, impact basin formation may predate the onset of the Martian core dynamo so that the magnetization of the southern highland terrain was caused by localized heating events followed by rapid cooling below the Curie temperature (Schubert *et al.*, 2000). The almost complete demagnetization of the Tharsis rise and other volcanic features implies that thermoremanent magnetization on Mars is confined to a relatively thin layer less than a few tens of kilometers thick. In a similar way, a possible former magnetization of the northern lowland crust may have been entirely erased by catastrophic volcanic flooding during the Hesperian period (Head *et al.*, 2002). Connerney *et al.* (2005) argue that variations in the crustal magnetic field can be associated with major geologic and topographic features, sometimes reminiscent of transform faults in oceanic crust on Earth. It is speculated that arcuate and linear magnetic features, similar to but much larger than those observed near mid-ocean ridges on Earth, could imply that early Noachian crust formation on Mars was accompanied by plate tectonics associated with crustal spreading in the

presence of a reversing dynamo (Connerney *et al.*, 2005).

Early attempts at modeling the deep interior structure of Mars suffered from poorly known values of its radius and moment of inertia. Improved measurements of the planet's mass M_p , radius r_p , gravitational potential and rotation rate by the Mariner, Viking, and Pathfinder spacecraft provided geodetic constraints required for models of the interior structure. The polar moment of inertia of Mars has been derived from a combined analysis of Mars Global Surveyor tracking and Mars Pathfinder and Viking Lander range and Doppler data. The re-analysis of the entire data set resulted in the most recent value of $C/M_p R_p^2 = 0.3650 \pm 0.0012$ (Yoder *et al.*, 2003), lower than the previously accepted value of $C/M_p R_p^2 = 0.3662 \pm 0.0017$ (Folkner *et al.*, 1997). The improved value is consistent with the model of a mostly hydrostatic planet with a non-hydrostatic contribution to the MoI factor entirely related to the axisymmetric distribution of topographic loads about Tharsis (Kaula, 1979). Using the improved value of $C/M_p R_p^2$ and taking into account the planet's gravitational oblateness and minor contributions due to the Tharsis rise results in a mean MoI factor of $I/M_p R_p^2 = 0.3635 \pm 0.0012$ and suggests a stronger concentration of mass toward the center than previously thought, with consequences for the planet's bulk chemistry and interior structure (Sohl *et al.*, 2005). As a consequence of the improved lower MoI factor, the Martian mantle may be less dense, about several tens of kg m^{-3} , with a smaller iron content than previously thought if crust thickness and core size are specified. It further implies that the Martian crust is several tens of kilometers thicker than previously thought if crust and mantle densities and core size are given. Finally, it suggests several tens of kilometers larger core radii if other parameters like core density, crust density, and crust thickness are fixed (Figure 14). If the crust thickness increases, dense mantle material will be replaced by less-dense crust material thereby reducing the planet's MoI factor. The mass deficit that arises cannot be compensated simply by increasing the core density, since this would even further reduce the MoI factor for a given core size. Therefore, to account for both the planet's mean density and mean MoI factor at constant core size, the mantle density will be required to increase and core density simultaneously to decrease. In terms of composition, the presence of a thicker crust requires the silicate mantle to be more enriched in iron with an increasing amount of light

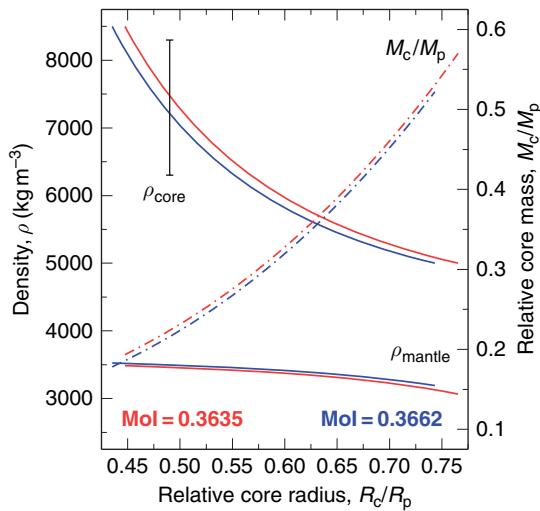


Figure 14 Two-layer model determinations of mantle and core density and core mass fraction M_c/M_p of Mars vs relative core radius R_c/R_p based on a Mol factor of 0.3635 (red) and 0.3662 (blue), respectively. Martian core densities are thought to range between those of γ -iron and iron sulfide as indicated by the error bar. Adapted from Sohl F, Schubert G, and Spohn T (2005). Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110: E12008 (doi:10.1029/2005JE002520).

alloying elements such as sulfur concentrated in the core. Additionally, the contribution of a thicker crust to the planet's bulk composition will become more pronounced (**Figure 15**).

The determination of the solar tidal potential Love number $k_2=0.153 \pm 0.017$, based on the analysis of 3 years of Mars Global Surveyor radio tracking data, indicates that the planet's interior is still sufficiently hot that at least the outer part of the Martian core is liquid (Yoder *et al.*, 2003). The reliable determination of the planet's tidal potential Love number will steadily improve as further spacecraft radio tracking data are collected, thereby additionally constraining the size of the Martian core. A small apparent tidal lag angle γ of about 0.7° has been obtained from a combined analysis of Viking and Mars Global Surveyor observations of Phobos' orbital position (Bills *et al.*, 2005). The corresponding estimate of Mars's tidal quality factor $Q_p=1/\tan \gamma \approx 85$ is within the range of former estimates of $Q_p=(100 \pm 50)$ (Smith and Born, 1976; Yoder, 1982). It has been shown previously that inelastic contributions to k_2 amount to only 1% of that caused by the planet's elastic tidal response if $Q_p \approx 100$ is assumed (Sohl and Spohn, 1997).

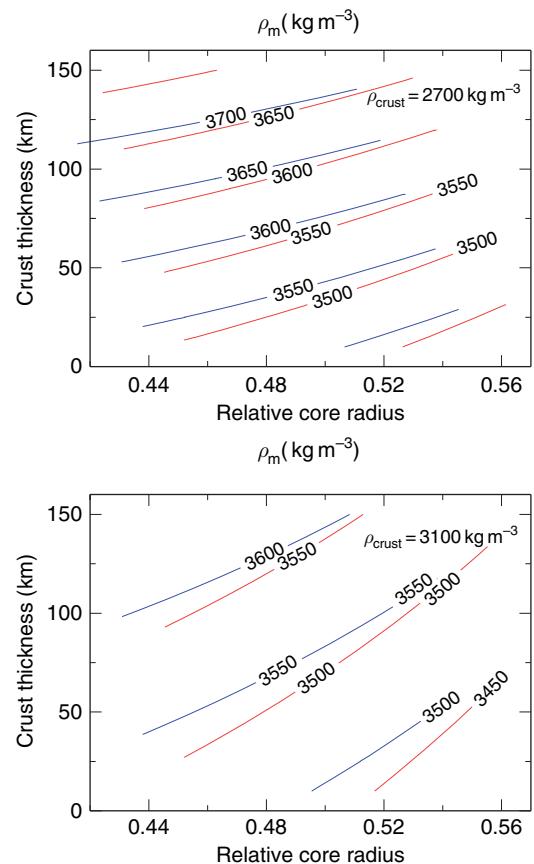


Figure 15 Contours of mantle density ρ_m as a function of relative core radius R_c/R_p and crust thickness based on a Mol factor of 0.3635 (red) and 0.3662 (blue), respectively. Mean crystal density is fixed at (top) 2700 and (bottom) 3100 kg m^{-3} . Adapted from Sohl F, Schubert G, and Spohn T (2005). Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110: E12008 (doi:10.1029/2005JE002520).

10.02.7.3 Composition

Considerations of element correlations in SNC meteorites and cosmochemical constraints suggest that Mars was formed by the homogeneous accretion of two geochemically distinct components, a highly reduced refractory component and a volatile-rich one. In this model, the former did not include elements more volatile than sodium or potassium, however, it contained all other elements in the same abundance ratios as in primitive carbonaceous chondrites of class CI. Iron and all siderophile elements are in the metallic state, and even silicon is partially metallic. This material existed mainly in the feeding zone of the growing Earth (Dreibus and Wänke, 1985; Wänke and Dreibus, 1988). In this

model, Mars is differentiated into a FeO-rich silicate mantle containing radiogenic heat sources in terrestrial abundances and a sulfur-rich Fe-Ni-FeS core of about 22% of the planet's mass containing about 14 wt.% sulfur with a radius roughly 50% of the surface radius. Longhi *et al.* (1992) have recast the mantle composition of Dreibus and Wänke (1985) into a pressure-dependent mineralogy resulting in an upper olivine-rich part, a transition zone composed of silicate-spinel, and a lower perovskite-rich layer. The model provides a mean dimensionless MoI factor of 0.353 and is consistent with the geochemical constraint of a bulk Fe/Si ratio of 1.71 representative of the composition of CI carbonaceous chondrites. From models that use identical sets of material parameter values for crust, mantle, and core, as derived from geochemical analyses of SNC meteorites, and that take into account self-compression and thermal expansion of the Martian interior, it has been concluded that it is impossible to reconcile the assumption of a bulk planet CI Fe/Si ratio with the observed value of the polar moment-of-inertia factor (Sohl and Spohn, 1997).

The Martian core is sufficiently large that even small changes in core size will result in significant changes of core volume. Since the core contributes most of the iron, a significant change in core volume can be expected to result in a notable modification of the bulk planet Fe/Si ratio. Modifications of the Fe/Si ratio due to changes in core volume may be partly compensated, however, by variable mantle iron contents in combination with crust thickness variations because of the significant mass fraction of the planet's silicate portion. The bulk-planet Fe/Si ratio is also notably dependent on the thermal state of the core because of the temperature dependence of the EOS parameters of iron and iron sulfide as the main core constituents (Fei *et al.*, 1995; Kavner *et al.*, 2001). In Figure 16, radially symmetric density models of the Martian interior are compared for variable sulfur content of the core. The sulfur contents of the Martian core is unknown, but geochemical models predict a sulfur-rich core (McSween, 1994). Furthermore, light constituents such as iron hydrite may also be present in the core depending on the amount of hydrogen dissolved in the core alloy when subject to the pressure and temperature conditions possibly prevailing in the Martian core (Gudkova and Zharkov, 2004). If core size and density are specified, high core temperatures may cause iron to be enriched in the core because thermal expansion of

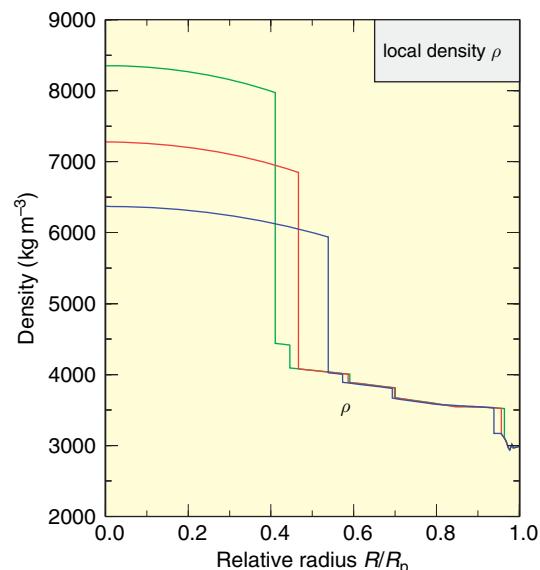


Figure 16 Radial density distribution of the Martian interior as a function of relative core radius R_c/R_p and core composition using the low-temperature crust model of Babeyko and Zharkov (2000). If the core were composed of pure iron (green), it would be small enough to enable the pressure-induced phase transformation from spinel to perovskite. In turn, a pure iron-sulfide core (blue) would be too large and mantle pressures, therefore, too low for the phase transition to occur. The intermediate density profile (red) refers to a Fe-FeS core containing iron and iron sulfide in equal amounts.

the main core constituents may exceed self-compression thereby increasing the planet's bulk Fe/Si ratio.

10.02.7.4 Mineralogy

Bertka and Fei (1997) obtained the sequence of mineralogical phase assemblages stable in the Martian mantle at elevated temperatures and up to core–mantle boundary pressures from high-pressure and high-temperature experiments using synthetic mineral mixtures according to the mantle composition of Dreibus and Wänke (1985). In Figure 17, it is seen that the upper part of the mantle then contains olivine, clinopyroxene, and garnet, whereas orthopyroxene is only present at pressures below 9 GPa. In the mantle transition zone at pressures above 13.5 GPa, β -spinel (wadsleyite) and clinopyroxene are subsequently replaced by γ -spinel (ringwoodite) and majorite until completion at about 17 GPa. A hot lower mantle is found to contain Mg-Fe silicate perovskite, magnesiowüstite, and majorite in the absence of CaSiO_3 perovskite (Bertka and Fei, 1997).

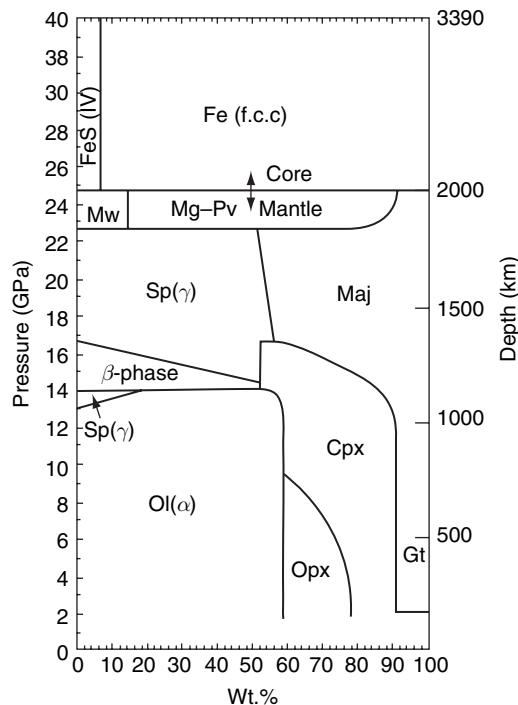


Figure 17 Mineral phase assemblages (in wt.%) of the Martian mantle as a function of pressure based on the chemical compositional model of Dreibus and Wänke (1985). Ol, olivine; Opx, orthopyroxene; Gt, garnet; Maj, majortite; Sp, spinel; Mw, magnesiowüstite; and Mg-Pv, Mg-Fe silicate perovskite. The location of the core–mantle boundary is from Fei et al. (1995). Adapted from Bertka CM and Fei Y (1997) Mineralogy of the Martian interior up to core–mantle boundary pressures. *Journal of Geophysical Research* 102: 5251–5264.

Verhoeven et al. (2005) have established empirical relationships between the most abundant mantle minerals based on a compilation of chemical compositions and related mineral assemblages of the silicate mantles of Mars and Earth. A perovskite layer at the base of the Martian mantle could only exist if pressures and temperatures in the mantle are sufficiently high for the occurrence of the mineral phase transitions. Therefore, the possible existence of a perovskite lower mantle and its stable phase assemblage in Mars not only depends on the core–mantle boundary pressure, that is, the size and composition of the core, but it is also sensitive to the temperature distribution deep inside the planet. van Thienen et al. (2006) have studied the stability range of such a putative perovskite lower mantle and its consequences for mantle plume dynamics, comparing the model mantle mineralogy from Bertka and Fei (1997) to

the model EH45 from Sanloup et al. (1999). Furthermore, uncertainties of a few 100 K in the experimental determinations of the pressure–temperature relation of the perovskite phase transformation and poor knowledge of the planet's thermal state (Breuer and Spohn, 2003) are taken into account. For a nominal core sulfur content of 14 wt.%, van Thienen et al. (2006) found that a perovskite layer sufficiently thick to affect mantle dynamics can be kept for both hot and cold mantle temperature profiles. Thus, it is feasible that at least in the early evolution of the planet, when mantle temperatures were much higher, a thin perovskite layer hovered above the core–mantle boundary. Additionally, data on the solid Fe–FeS system at high pressure and high temperature show a high-pressure, high-density phase change to a hexagonal NiAs superstructure at core pressures (Fei et al., 1995). This phase transition together with the assumed sulfur content in the Martian core of 12–16% of the core mass derived from the chemistry of the SNC meteorites also would imply a smaller core size and, in turn, an increase in thickness of a tentative perovskite lower mantle. Alternatively, a smaller size of the Martian core in favor of a deep perovskite layer is conceivable if a solid inner core composed of γ-iron would be present and surrounded by a less dense, volatile-rich liquid outer core.

If the geochemical requirement of a CI chondrite bulk composition is kept, the mantle density profile obtained from the experimentally determined sequence of mineralogic phase assemblages (Bertka and Fei, 1997) and a core density consistent with a core sulfur content of 14 wt.% would result in a mean moment of inertia factor of 0.354 (Bertka and Fei, 1998a). However, this model which lacks a perovskite layer at the base of the mantle requires a crust thickness of 180–320 km assuming a crustal density of $2700\text{--}3000 \text{ kg m}^{-3}$. Since such a crust thickness is considered to be unrealistically large, a CI chondrite bulk composition of the Martian interior as assumed by Dreibus and Wänke (1985) may be questionable (Bertka and Fei, 1998b). Models using the same mantle density profile and a range of model core compositions but allowing for the moment of inertia factor of 0.3662 as obtained from the Pathfinder measurements generally produce bulk Fe/Si ratios below the CI chondrite value of 1.71. This may suggest that the formation of Mars and the terrestrial planets cannot be explained solely by the accretion

of CI carbonaceous chondrite material (Bertka and Fei, 1998b).

10.02.7.5 Martian Seismicity

A three-axis short-period seismometer onboard the Viking Lander 2 at Utopia Planitia collected more than 600 h of high-quality data to explore the seismic environment of Mars (Anderson *et al.*, 1976, 1977). The seismometer was added late in the mission planning, and therefore had to be installed on the top of the lander structure. Hence, wind noise contaminated the seismic data. No large seismic events were detected, indicating that Mars is less seismically active than Earth by an estimated three orders of magnitude. A seismic signal from one small quake was possibly recorded. Its tentative analysis suggested that it had a magnitude of 3 and had occurred at a distance of 110 km. Identification of late arrivals implied the presence of a crust with a thickness of 15 km near the landing site. However, it cannot be ruled out that wind gusts mimicked a seismic event.

While the seismicity of Mars is largely unexplored, there are indications that Mars could be quite active. Most morphological features on Mars seem to be only partially isostatically compensated

(Zuber *et al.*, 2000). These include the Tharsis rise as well as the Chryse and Amazonis basins. Regions of partial compensation on Earth are generally the most seismically active. Tectonic stresses implied by a lack of compensation are known to be in the 10- to 100-bar range and stress drops in earthquakes are in the same range. Hence, Mars could be tectonically active, though seismicity may be more localized. Based on a global compilation of compressional and extensional faults derived from MOLA shaded relief data, Knapmeyer *et al.* (2006) have constructed a new Martian seismicity model that predicts up to 25 seismic events per year with moment magnitudes greater than 4 corresponding to seismic moments $\geq 1.26 \times 10^{15}$ N m. Most of these events are expected to occur near the Tharsis rise but other seismic centers may be located south of Hellas and north of Utopia Planitia (Figure 18).

10.02.8 Venus

Venus, with a radius of 6051.8 km, is only slightly smaller than Earth, but the small size difference may have important consequences for the planet's interior.

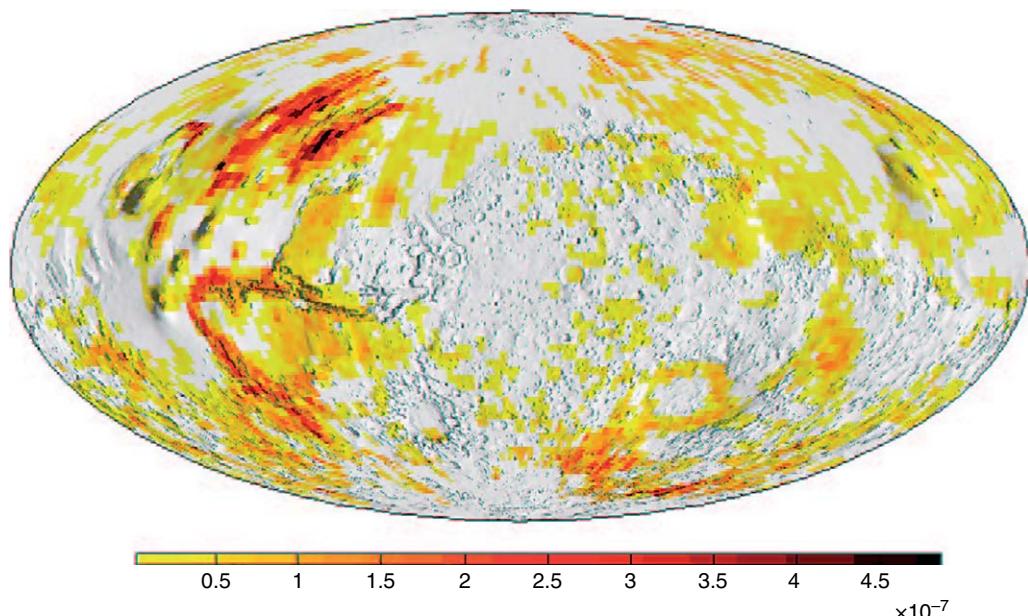


Figure 18 Global distribution of quake probability based on the distribution of extensional and compressional faults on Mars in units of faults per km^2 , irrespective of geologic age. Adapted from Knapmeyer M, Oberst J, Hauber E, Wählisch M, Deuchler C, and Wagner R (2006) Working models for spatial distribution and level of Mars' seismicity. *Journal of Geophysical Research* 111: E11006 (doi:10.1029/2006JE002708).

10.02.8.1 General

Unlike Earth, Venus has no magnetic field (Russell, 1980; Phillips and Russell, 1987; Donahue and Russell, 1997). This is likely a consequence of the smaller pressure at the center of Venus compared with the pressure at the center of Earth's core. Because of the lower pressure, it is possible that Venus' core has not yet cooled sufficiently to initiate inner-core growth, but has cooled enough to prevent the operation of a purely thermally driven dynamo (Stevenson *et al.*, 1983). Venus' lack of a magnetic field could also be due to its lack of plate tectonics, perhaps indicative of a sluggish form of mantle convection that is unable to cool the core efficiently enough to initiate thermal dynamo action (Nimmo and Stevenson, 2000). Still another possibility, though probably unlikely, is that the core of Venus has solidified enough that a dynamo cannot operate in the remaining liquid outer shell (Arkani-Hamed, 1994). Another consequence of Venus' slightly smaller size compared with Earth is that the perovskite–post-perovskite phase transition that occurs near the base of the Earth's mantle may not occur in Venus' mantle. If the core were to contain less light elements than the Earth's core, however, the Venusian mantle could be even deeper than that of the Earth.

10.02.8.2 Interior Structure

The above discussion tacitly assumed that the structure of Venus' interior is similar to that of the Earth, that is, that Venus is basically a three-layer body with a metallic core surrounded by a rocky mantle which is in turn surrounded by a compositionally distinct rocky crust. The basic structure of Venus is illustrated in **Figure 19**. However, the structure of the Venusian interior is a matter of some guesswork because we do not know the MoI of the planet and though spacecraft have visited Venus and landed on its surface, we have not yet seismically probed its interior. Because of the high pressure (≈ 95 bars) and temperature (≈ 737 K) at the surface of Venus (Seiff, 1983), a seismic experiment is probably far into the future. Venus' MoI is also something we will not know anytime soon. Venus lacks a moon to force a precession of the planet's retrograde spin axis, but the torque on the solar-induced tide forces a free precession ranging from 44.1 to $45.6'' \text{ yr}^{-1}$ and a variation of obliquity, or wobble, estimated at about $1.1'' \text{ yr}^{-1}$ (Yoder, 1997). Nevertheless, Venus is such a large terrestrial planet and would be heated so thoroughly

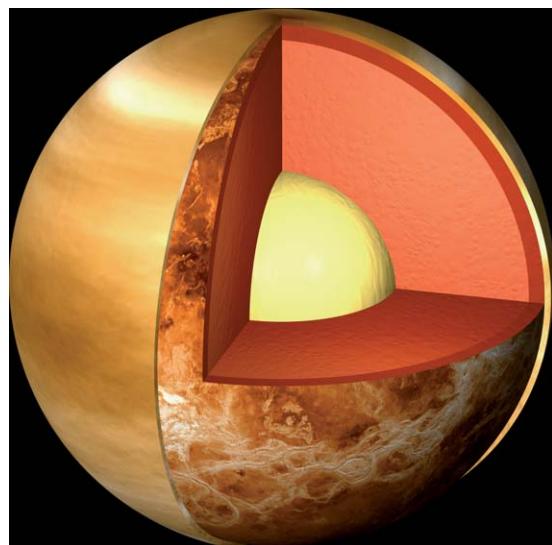


Figure 19 Cut-away view of the interior of Venus.
© Calvin J. Hamilton.

upon accretion, that differentiation into an Earth-like structure is all but inevitable. The extensive basaltic plains on Venus (Surkov, 1983; Weitz and Basilevsky, 1993; Basilevsky *et al.*, 1992; Campbell *et al.*, 1997) are evidence that it has differentiated a crust; it is even possible that plateau-like highlands such as Alpha Regio represent compositionally distinct pieces of crust.

Doppler radio tracking data of the Magellan and Pioneer Venus spacecraft have provided basic information about the planet's mass ($G M = 324858.6 \pm 0.014 \text{ km}^3 \text{ s}^{-2}$, G is the gravitational constant), gravitational field, and tidal Love number k_2 (Sjogren *et al.*, 1997). The Love number k_2 is determined from the time variations in the gravitational coefficients C_{22} and S_{22} at the solar period. Konopliv and Yoder (1996) have found $k_2 = 0.295 \pm 0.066$, a value consistent with a liquid core (Yoder, 1997). A 180 degree and order spherical harmonic model of Venus' gravity is given in Konopliv *et al.* (1999). Magellan radar altimetry data have been used to produce the 360 degree and order spherical harmonic model of Venus' topography presented in Rappaport *et al.* (1999) (see Chapter 10.5).

10.02.8.3 Composition

The predominantly basaltic nature of Venus' surface is suggested by the geochemical data obtained by the Venera and Vega landers and the morphology of the

widespread volcanic landforms (Grimm and Hess, 1997). The K/U and K/Th ratios of Venusian samples are similar to those obtained for terrestrial volcanics, SNC meteorites, and Martian samples. The variation of the K/U ratios at the seven landing sites is relatively narrow within a factor of three of each other implying that the volatile-refractory element inventory of Venus is comparable to that of Earth and Mars, but different from that of the Moon. The major oxide compositions of Venusian rocks are broadly consistent with those of basaltic rocks so that it is reasonable to assume that the crust is largely basaltic (Grimm and Hess, 1997). The gravity and topography data can be analyzed to provide information on the average thickness of the crust; estimates lie in the range 20–50 km (Grimm and Hess, 1997; Nimmo and McKenzie, 1998). The crust is thicker (45–85 km) beneath the plateau highlands (Alpha, Ovda, Thetis, and Tellus Regiones) (Moore and Schubert, 1997) compared with other regions. The basalt–eclogite phase change limits the crustal thickness to about 50 km for moderate conductive geotherms of $>5\text{--}10 \text{ K km}^{-1}$ (Grimm and Hess, 1997). The Venusian gravity and topography data have also been used to infer values for the thickness of the planet's thermal lithosphere and its elastic upper layer. Estimates of lithosphere thickness vary between about 200 and 400 km with thinner lithosphere beneath volcanic highlands (e.g., Atla and Beta Regiones) (Moore and Schubert, 1997; Kucinskas and Turcotte, 1994; Phillips, 1994; Smrekar, 1994; Simons *et al.*, 1994; Herrick *et al.*, 1989). Lithospheric thinning beneath the volcanic rises could have been caused by mantle plumes. There is buoyant sublithospheric mantle beneath the volcanic rises (Moore and Schubert, 1997). The top part of the lithosphere that behaves elastically is only about 30 km thick (McKenzie and Nimmo, 1997). Unlike on Earth, gravity anomalies correlate with high topography on Venus.

10.02.8.4 Tectonism

Pioneer Venus radar, Earth-based radar observations, Venera 15–16 orbital imaging radar, and Magellan radar images have provided views of the surface of Venus unimpeded by the global cloud cover that prevents visual observation. These views, together with the topography and gravity data, reveal the nature of Venusian tectonism and volcanism. On Earth, the global oceanic rift system and the arcuate ocean trenches are the primary surface

manifestations of plate tectonics. The almost total absence of these features on Venus has led to the conclusion that active plate tectonics is not occurring on Venus (Kaula and Phillips, 1981; Kaula, 1994). At the present time Venus is a one-plate planet. Nevertheless, there are tectonic features on Venus that resemble major tectonic features on Earth. Beta Regio, a volcanic highland, has many of the features of a continental rift on Earth. It has a domal structure with a diameter of about 2000 km and a swell amplitude of about 2 km. It has a well-defined central rift valley with a depth of 1–2 km and there is some evidence for a three-armed planform (aulacogen). Alta, Eistla, and Bell Regiones have similar rift zone characteristics (Grimm and Phillips, 1992; Senske *et al.*, 1992). Aphrodite Terra with a length of some 1500 km is reminiscent of major continental collision zones on Earth, such as the mountain belt that extends from the Alps to the Himalayas. Ishtar Terra is a region of elevated topography with a horizontal scale of 2000–3000 km. A major feature is Lakshmi Planum, which is an elevated plateau similar to Tibet with a mean elevation of about 4 km. This plateau is surrounded by linear mountain belts, Akna, Danu, Freyja, and Maxwell montes, reaching elevations of 10 km, similar in scale and elevation to the Himalayas (Kaula *et al.*, 1997).

10.02.8.5 Dynamics

The impact crater population on the surface of Venus has been used to infer a mean surface age of several hundred to as much as 800 Myr (McKinnon *et al.*, 1997; Herrick *et al.*, 1997). It has been proposed that the relatively young age of Venus' surface was set in a global volcanic resurfacing event and that relatively little volcanism has occurred since (Schaber *et al.*, 1992; Basilevsky *et al.*, 1997). The resurfacing event could be the means by which Venus expels its heat. One way this could happen is the global foundering of a thick, relatively cold and heavy lithosphere and its replacement by the relatively hot underlying mantle (Turcotte, 1993). Such events might have occurred episodically throughout Venus' history. Between such events the lithosphere would thicken but Venus would have no efficient way, like plate tectonics on Earth, to expel its heat. Instead the heat building up in the interior during the quiescent period would be lost in the mantle overturn when the lithosphere thickened enough to become gravitationally unstable. The initiation of such an event might be evident today on Venus' surface in the form of

large coronae. Coronae are quasi-circular features, 100–2600 km in diameter, with raised interiors and elevated rims, often with annular troughs (Stofan *et al.*, 1997). Mckenzie *et al.* (1992) and Sandwell and Schubert (1992a, 1992b) have argued that the perimeters of several large coronae on Venus, specifically Artemis, Latona, and Eithinoa, resemble terrestrial subduction zones in both planform and topography. Artemis chasma has a radius of curvature similar to that of the South Sandwich subduction zone on the Earth. Sandwell and Schubert (1992a) proposed that the large coronae are incipient circular subduction zones. The foundering lithosphere is replaced by ascending hot mantle in a manner similar to back-arc spreading on the Earth. A single global resurfacing event for Venus has been challenged by Hauck *et al.* (1998). They argue that the interpretation of the Venusian crater distribution is nonunique and they identify some units in the volcanic plains that have a spread in age of about 0.5 Gyr.

Because Venus lacks plate tectonics, convection in its mantle is different from the style of convection in Earth's mantle. Venusian mantle convection occurs in the sluggish or stagnant lid regime, that is, it is confined below the lithosphere or nearly rigid lid (Schubert *et al.*, 1997). This form of convection is less efficient at transporting heat than is the plate tectonic regime with consequent implications for the thermal history of the planet and the dynamics of its core, as noted above. Mantle convection in Venus may be unable to establish a near equilibrium with its internal heat sources resulting in the episodic overturning of its mantle and global resurfacing. Enhanced core cooling would occur during such an event with the possible initiation of a transient dynamo. Heat from the core would be carried away by mantle plumes that could form volcanic rises similar to Atla and Beta Regiones.

10.02.9 Summary and Outlook

The terrestrial planets Mercury, Venus, Earth, and Mars have low masses, small radii, and large densities in comparison to the giant planets in the outer solar system. This is also true for terrestrial-type bodies like the Moon and some of the outer planet satellites and it provides important clues on their bulk compositions. Rotational, gravitational, and magnetic field observations indicate that the interiors of these bodies are strongly differentiated and subdivided like that of the Earth into iron-rich cores, silicate

mantles, and rocky crusts derived from partial mantle melts. Isotope data reveal that the cores and the primary crusts formed early and rapidly. Geodetic observations of the rotational state and/or tidal response suggest that the interiors are warm enough to maintain liquid outer core shells or entirely liquid cores. For Mars, Venus, and Earth, mantle pressures are sufficient to permit mineral phase transformations from olivine and pyroxene assemblages to spinel or even perovskite and post-perovskite phases. Since the phase transition depths also depend on the ambient temperature and the iron content of the mantle rocks, future seismological observations have the potential to provide additional information on the thermal states and compositional differences of the terrestrial planets. Single-plate planets, the Moon, Mercury, Mars, and Venus, are believed to be cooling by lithospheric thickening while the deep interior remains relatively warm. It is likely, therefore, that due to the progressive cooling of the planet's outer portion, thermoelastic stresses will be occasionally released at preexisting faults thereby causing local seismic activity at a level detectable by seismometers.

Future developments will benefit from the combination of classic forward modeling techniques and Bayesian probabilistic inversion methods as more spacecraft and laboratory data are being collected that considerably extend the database required to construct planetary interior models. The inverse problem is commonly based on nonlinear relations between model and data. Uncertainties of observed and calculated parameters have to be taken into account, as well as *a priori* knowledge such as the possible range of surface composition as obtained from remote-sensing observations and *in situ* measurements. This has been detailed by Khan *et al.* (2007) who have directly inverted geophysical observations – including the lunar mass and MoI, the Apollo seismic data set, and long-period electromagnetic sounding data – for the chemical composition, mantle mineralogy, and thermal state of the Moon. A Markov chain Monte Carlo algorithm is applied to perform the joint inversion of these data sets. The calculation of mineral phase equilibria in the CFMAS system comprising rock-forming oxides CaO–FeO–MgO–Al₂O₃–SiO₂ is based on Gibbs energy minimization schemes (Kuskov and Galimzyanov, 1986).

Future space telescope missions like *Kepler* and *COROT* are expected to detect many super-massive terrestrial-type exoplanets in the coming years. From

combined photometric transit observations and Doppler radial velocity measurements, mass and radius of a transiting exoplanet will be obtained and reliable estimates of its mean density derived (see Chapter 10.13) It is necessary, therefore, to establish mass-radius relations of massive terrestrial-type bodies for comparison with these new observational data. The mass-radius relation of these objects is likely to deviate from a cubic power law for constant-density scaling due to large self-compression and higher internal temperatures. Valencia *et al.* (2006) have developed scaling laws for the interiors of super-massive terrestrial exoplanets yielding their total radius, mantle thickness, core size, and mean density as a function of total mass. For Super-Earths, a relation between total radius R and mass M like $R \propto M^{0.267-0.272}$ has been suggested. Furthermore, it is seen that uncertainties that are linked to thermodynamic EOS parameters do not significantly alter the scaling exponents of bulk properties at a given pressure regime. In contrast, the scaling exponents are mainly affected by compositional differences (super-massive rock, metal, ice/water bodies). This implies that the bulk compositions of transiting super-massive terrestrial-type exoplanets could be measured soon by using Earth-orbiting space telescopes (Valencia *et al.*, 2007).

Acknowledgments

The authors thank Tamara Gudkova, Martin Knapmeyer, and Jürgen Oberst for helpful discussions and acknowledge constructive criticisms from Tim Van Hoolst, Mark Wieczorek, and Tilman Spohn. Some figures in this chapter were produced by using the Generic Mapping Tools (GMT) of Wessel and Smith (1991).

References

- Acuña MH, Connerney JEP, Wasilewski P, *et al.* (1998) Magnetic field and plasma observations at Mars: Initial results of the Mars global surveyor mission. *Science* 279: 1676–1680.
- Acuña MH, Connerney JEP, Ness NF, *et al.* (1999) Global distribution of crustal magnetization discovered by the Mars global surveyor MAG/ER experiment. *Science* 284: 790–793.
- Anderson DL (1989) *Theory of the Earth*. Cambridge, MA: Blackwell.
- Anderson DL, Miller WF, Latham GV, *et al.* (1976) The Viking seismic experiment. *Science* 194: 1317–1321.
- Anderson DL, Miller WF, Latham GV, *et al.* (1977) Seismology on Mars. *Journal of Geophysical Research* 82: 4524–4546.
- Anderson JD, Colombo G, Esposito PB, Lau EL, and Trager GB (1987) The mass, gravity field, and ephemeris of Mercury. *Icarus* 71: 337–349.
- Anderson JD, Jurgens RF, Lau EL, Slade MA, III, and Schubert G (1996) Shape and orientation of Mercury from radar ranging data. *Icarus* 124: 690–697.
- Anderson JD and Schubert G (2007) Saturn's satellite Rhea is a homogeneous mix of rock and ice. *Geophysical Research Letters* 34: L02202.
- Anderson OL (1984) A universal thermal equation-of-state. *Journal of Geodynamics* 1: 185–214.
- Anderson OL, Isaak DG, and Oda H (1992) High-temperature elastic constant data on minerals relevant to geophysics. *Reviews of Geophysics* 30: 57–90.
- Arkani-Hamed J (1994) On the thermal evolution of Venus. *Journal of Geophysical Research* 99: 2019–2033.
- Babeyko AY and Zharkov VN (2000) Martian crust: A modeling approach. *Physics of the Earth and Planetary Interiors* 117: 421–435.
- Badding JV, Hemley RJ, and Mao HK (1991) High-pressure chemistry of hydrogen in metals: *In situ* study of iron hydride. *Science* 253: 421–424.
- Balog PS, Secco RA, Rubie DC, and Frost DJ (2003) Equation of state of liquid Fe-10 wt% S: Implications for the metallic cores of planetary bodies. *Journal of Geophysical Research* 108(B2): 2124 (doi:10.1029/2001JB001646).
- Balogh A and Giampieri G (2002) Mercury: The planet and its orbit. *Reports of Progress in Physics* 65: 529–560.
- Banks R (1969) Geomagnetic variations and the electrical conductivity of the upper mantle. *Geophysical Journal of the Royal Astronomical Society* 17: 457–487.
- Basilevsky AT, Nikolaeva OV, and Weitz CM (1992) Geology of the Venera-8 landing site region from Magellan data – Morphological and geochemical considerations. *Journal of Geophysical Research* 97(E10): 16,315–16,335.
- Basilevsky AT, Head JW, Schaber GG, and Strom RG (1997) The resurfacing history of Venus. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1047–1084. Tucson, AZ: University of Arizona Press.
- Bass J (2004) *Current and Future Research Directions in High-Pressure Mineral Physics*. Consortium for Materials Properties Research in Earth Sciences (COMPRES), Stony Brook, NY.
- Belleguic V, Lognonné P, and Wieczorek M (2005) Constraints on the Martian lithosphere from gravity and topography data. *Journal of Geophysical Research* 110: E11005 (doi:10.1029/2005JE002437).
- Bertka CM and Fei Y (1997) Mineralogy of the Martian interior up to core–mantle boundary pressures. *Journal of Geophysical Research* 102: 5251–5264.
- Bertka CM and Fei Y (1998a) Density profile of an SNC model Martian interior and the moment of inertia factor of Mars. *Earth and Planetary Science Letters* 157: 79–88.
- Bertka CM and Fei Y (1998b) Implications of Mars Pathfinder data for the accretion history of the terrestrial planets. *Science* 281: 1838–1840.
- Bills BG and Ferrari AJ (1977) A lunar density model consistent with topographic, gravitational, librational, and seismic data. *Journal of Geophysical Research* 82: 1306–1314.
- Bills BG, Neumann GA, Smith DE, and Zuber MT (2005) Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110: E07004 (doi:10.1029/2004JE002376).
- Bina CR and Helffrich GR (1992) Calculation of elastic properties from thermodynamic equation of state principles. *Annual Review of Earth and Planetary Sciences* 20: 527–552.

- Binder AB (1998) Lunar prospector: Overview. *Science* 281: 1475–1476.
- Blewett DT, Lucey PG, Hawke BR, Ling GG, and Robinson MS (1997) A comparison of Mercurian reflectance and spectral quantities with those of the Moon. *Icarus* 129: 217–231.
- Boehler R (1992) Melting of the Fe–FeO and the Fe–FeS systems at high pressure: Constraints on core temperatures. *Earth and Planetary Science Letters* 111: 217–227.
- Boehler R (1996a) Fe–FeS eutectic temperatures to 620 kbar. *Physics of the Earth and Planetary Interiors* 96: 181–186.
- Boehler R (1996b) Melting of mantle and core materials at very high pressures. *Philosophical Transactions of the Royal Society of London A* 354: 1265–1278.
- Boehler R (1996c) Melting temperature of the Earth's mantle and core: Earth's thermal structure. *Annual Review of Earth and Planetary Sciences* 24: 15–40.
- Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics on Mars: Evidence for magnetic field history and crust evolution. *Journal of Geophysical Research* 108(E7): 5072 (doi:10.1029/2002JE001999).
- Britt DT and Consolmagno GJ (2003) Stony meteorite porosities and densities: A review of the data through 2001. *Meteorites and Planetary Science* 38: 1161–1180.
- Brückner J, Dreibus G, Rieder G, and Wänke H (2003) Refined data of Alpha-Proton-X-ray spectrometer analyses of soils and rocks at the Mars Pathfinder site: Implications for surface chemistry. *Journal of Geophysical Research* 108(E12): 8094 (doi:10.1029/2003JE002060).
- Burns JA (1976) Consequences of the tidal slowing of Mercury. *Icarus* 28: 453–458.
- Bussey DBJ and Spudis PD (2000) Compositional studies of the Orientale, Humorum, Nectaris, and Crisium lunar basins. *Journal of Geophysical Research* 105: 4235–4244.
- Cameron AGW, Fegley B, Benz W, and Slattery WL (1988) The strange density of Mercury: Theoretical considerations. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 692–708. Tucson, AZ: University of Arizona Press.
- Campbell BA, Arvidson RE, Shepard MK, and Brackett RA (1997) Remote sensing of surface processes. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 503–526. Tucson, AZ: University of Arizona Press.
- Chapman S and Price AT (1930) The electric and magnetic state of the interior of the earth, as inferred from terrestrial magnetic variations. *Philosophical Transactions of the Royal Society of London Series A* 229: 0427–0460.
- Chase SC, Miner ED, Morrison D, Münch G, and Neugebauer G (1976) Mariner 10 infrared radiometer results: Temperatures and thermal properties of the surface of Mercury. *Icarus* 28: 565–578.
- Chenet H, Lognonné P, Wieczorek M, and Mizutani H (2006) Lateral variations of lunar crustal thickness from the Apollo seismic data set. *Earth and Planetary Science Letters* 243: 1–14.
- Connerney JEP, Acuña MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Review* 111: 1–32.
- Connerney JEP, Acuña MH, Ness NF, et al. (2005) Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences of the United States of America* 102: 14,970–14,975.
- Connerney JEP, Acuña MH, Wasilewski PJ, et al. (1999) Magnetic lineations in the ancient crust of Mars. *Science* 284: 794–798.
- Connerney JEP, Acuña MH, Wasilewski PJ, et al. (2001) The global magnetic field of Mars and implications for crustal evolution. *Geophysical Research Letters* 28: 4015–4018.
- Constable S and Constable C (2004) Observing geomagnetic induction in magnetic satellite measurements and associated implications for mantle conductivity. *Geochemistry, Geophysics, Geosystems* 5: Q01006.
- Conzelmann V (1994) *Thermische Evolution des Planeten Merkur berechnet unter Anwendung verschiedener Viskositätsgesetze*. PhD Thesis, Institut für Planetologie, Westfälische Wilhelms-Universität, Münster.
- Dehant V, Defraigne P, and Hoolst TV (2000) Computation of Mars' transfer function for nutations, tides, and surface loading. *Physics of the Earth and Planetary Interiors* 117: 385–395.
- Dickey JO, Bender PL, Faller JE, et al. (1994) Lunar laser ranging: A continuing legacy of the Apollo program. *Science* 265: 482–490.
- Donahue TM and Russell CT (1997) The Venus atmosphere and ionosphere and their interaction with the solar wind: An overview. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 3–31. Tucson, AZ: University of Arizona Press.
- Drake MJ (2001) The eucrete/Vesta story. *Meteoritics* 36: 501–513.
- Dreibus G and Wänke H (1985) Mars: A volatile rich planet. *Meteoritics* 20: 367–382.
- Elkins-Tanton LT, Chatterjee N, and Grove TL (2003a) Experimental and petrologic constraints on lunar differentiation from the Apollo 15 green picritic glasses. *Meteorites and Planetary Science* 38: 515–527.
- Elkins-Tanton LT, Parmentier EM, and Hess PC (2003b) Magma ocean fractional crystallization and cumulate overturn in terrestrial planets: Implications for Mars. *Meteorites and Planetary Science* 38: 1753–1771.
- Elphic RC and Russell CT (1978) On the apparent source depth of planetary magnetic fields. *Geophysical Research Letters* 5: 211–214.
- Esposito PB, Banerdt WB, Lindal GF, et al. (1992) Gravity and topography. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 209–248. Tucson, AZ: University of Arizona Press.
- Fei Y, Prewitt CT, Mao HK, and Bertka CM (1995) Structure and density of FeS at high pressure and high temperature and the internal structure of Mars. *Science* 268: 1892–1894.
- Fei Y, Bertka CM, and Finger LW (1997) High-pressure iron-sulfur compound, Fe_3S_2 , and melting relations in the Fe–FeS system. *Science* 268: 1621–1623.
- Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars Pathfinder. *Science* 278: 1749–1752.
- Frey HV, Roark JH, Shockley KM, Frey EL, and Sakimoto SEH (2002) Ancient lowlands on Mars. *Geophysical Research Letters* 29(10): 1384 (doi:10.1029/2001GL013832).
- Ghosh A and McSween HJ, Jr. (1998) A thermal model for the differentiation of asteroid 4 Vesta, based on radiogenic heating. *Icarus* 134: 187–206.
- Goettel KA (1988) Present bounds on the bulk composition of Mercury: Implications for planetary formation processes. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 613–621. Tucson, AZ: University of Arizona Press.
- Goins NR, Dainty AM, and Toksöz MN (1981) Lunar seismology: The internal structure of the Moon. *Journal of Geophysical Research* 86: 5061–5074.
- Grasset O, Sotin C, and Deschamps F (2000) On the internal structure and dynamics of Titan. *Planetary and Space Sciences* 48: 617–636.
- Grimm RE and Hess PC (1997) The crust of Venus. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 3–31. Tucson, AZ: University of Arizona Press.

- Environment*, pp. 1205–1244. Tucson, AZ: University of Arizona Press.
- Grimm RE and Phillips RJ (1992) Anatomy of a Venusian hot-spot – Geology, gravity, and mantle dynamics of Eistla Regio. *Journal of Geophysical Research* 97(E10): 378–388.
- Gudkova TV and Zharkov VN (2004) Mars: Interior structure and excitation of free oscillations. *Physics of the Earth and Planetary Interiors* 142: 1–22.
- Hale AS and Hapke B (2002) A time-dependent model of radiative and conductive thermal energy transport in planetary regoliths with applications to the Moon and Mercury. *Icarus* 156: 318–334.
- Halekas JS, Mitchell DL, Lin RP, et al. (2001) Mapping of crustal magnetic anomalies on the lunar near side by the Lunar Prospector electron reflectometer. *Journal of Geophysical Research* 106(E11): 27841–27852.
- Harder H and Schubert G (2001) Sulfur in Mercury's core? *Icarus* 151: 118–122.
- Harmon JK, Perillat PJ, and Slade MA (2001) High-resolution radar imaging of Mercury's north pole. *Icarus* 149: 1–15.
- Hauck SA, Dombard AJ, Phillips RJ, and Solomon SC (2004) Internal and tectonic evolution of Mercury. *Earth and Planetary Science Letters* 222: 713–728.
- Hauck SA, Phillips RJ, and Price MH (1998) Venus: Crater distribution and plains resurfacing models. *Journal of Geophysical Research* 103(E6): 13635–13642.
- Head JW, Kreslavsky MA, and Pratt S (2002) Northern lowlands of Mars: Evidence for widespread volcanic flooding and tectonic deformation in the Hesperian Period. *Journal of Geophysical Research* 107(E1): 5003 (doi:10.1029/2001JE001445).
- Heiken G, Vaniman D, and French BM (1991) *The Lunar Sourcebook*. New York: Cambridge University Press.
- Hemley RJ (2006) Erskine Williamson, extreme conditions, and the birth of mineral physics. *Physics Today* 59: 50–56.
- Hemley RJ and Ashcroft NW (1998) The revealing role of pressure in the condensed matter sciences. *Physics Today* 51: 26–32.
- Herrick RR, Bills BG, and Hall SA (1989) Variations in effective compensation depth across Aphrodite Terra, Venus. *Geophysical Research Letters* 16: 543–546.
- Herrick RR, Sharpton VL, Malin MC, Lyons SN, and Feely K (1997) Morphology and morphometry of impact craters. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1015–1046. Tucson, AZ: University of Arizona Press.
- Hiesinger H, Head JW, Wolf U, Jaumann R, and Neukum G (2003) Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum. *Journal of Geophysical Research* 108(E7): 5065 (doi:10.1029/2002JE001985).
- Hobbs BA (1987) Conductivity profiles from global data. *Pure and Applied Geophysics* 125: 393–407.
- Hofmeister AM (1991) Pressure derivatives of the bulk modulus. *Journal of Geophysical Research* 96: 21893–21907.
- Hood LL (1986) Geophysical constraints on the lunar interior. In: Phillips RJ, Hartmann WK, and Taylor GJ (eds.) *Origin of the Moon*, pp. 397–409. Houston: Lunar and Planetary Institute.
- Hood LL, Herbert F, and Sonett CP (1982) The deep lunar electrical conductivity profile – Structural and thermal inferences. *Journal of Geophysical Research* 87: 5311–5326.
- Hood LL and Huang Z (1991) Formation of magnetic anomalies antipodal to lunar impact basins: Two-dimensional model calculations. *Journal of Geophysical Research* 96: 9837–9846.
- Hood LL and Jones JH (1987) Geophysical constraints on lunar bulk composition and structure: A reassessment. *Proceedings of the 17th Lunar and Planetary Science Conference Part 2, Journal of Geophysical Research* 92: E396–E410.
- Hood LL, Mitchell DL, Lin RP, Acuna MH, and Binder AB (1999) Initial measurements of the lunar induced magnetic dipole moment using Lunar Prospector magnetometer data. *Geophysical Research Letters* 26: 2327–2330.
- Hood LL and Sonett CP (1982) Limits on the lunar temperature profile. *Geophysical Research Letters* 9: 37–40.
- Hood LL, Zakharian A, Halekas J, et al. (2001) Initial mapping and interpretation of lunar crustal magnetic anomalies using Lunar Prospector magnetometer data. *Journal of Geophysical Research* 106(E11): 27825–27839.
- Hood LL and Zuber MT (2000) Recent refinements in geophysical constraints on lunar origin and evolution. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 397–409. Tucson, AZ: University of Arizona Press.
- Hussmann H, Sohl F, and Spohn T (2006) Subsurface oceans and deep interiors of medium-sized outer planet satellites and large trans-neptunian objects. *Icarus* 185: 258–273.
- Jarvis GT and Peltier WR (1989) Convection models and geophysical observations. In: Peltier WR (ed.) *Mantle Convection: Plate Tectonics and Global Dynamics*, pp. 479–593. New York: Gordon and Breach.
- Jeanloz R, Mitchell DL, Sprague AL, and de Pater I (1995) Evidence for a basalt-free surface on Mercury and implications for internal heat. *Science* 268: 1455–1457.
- Jehn R, Corral C, and Giampieri G (2004) Estimating Mercury's 88-day libration amplitude from orbit. *Planetary and Space Sciences* 52: 727–732.
- Jolliff BL, Gillis JJ, Haskin LA, Korotev RL, and Wieczorek MA (2000a) Major lunar crustal terranes: Surface expressions and crust-mantle origins. *Journal of Geophysical Research* 105: 4197–4216.
- Jolliff BL, Gaddis LR, Ryder G, et al. (2000b) New views of the Moon: Improved understanding through data integration. *EOS Transactions of the American Geophysical Union* 81(349): 354–355.
- Kaula WM (1979) The moment of inertia of Mars. *Geophysical Research Letters* 6: 194–196.
- Kaula WM (1994) The tectonics of Venus. *Philosophical Transactions of the Royal Society of London, Series A* 349: 345–355.
- Kaula WM, Drake MJ, and Head JW (1986) The Moon. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 581–628. Tucson, AZ: University of Arizona Press.
- Kaula WM, Lenardic A, Bindschadler DL, and Arkani-Hamed J (1997) Ishtar terra. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 879–900. Tucson, AZ: University of Arizona Press.
- Kaula WM and Phillips RJ (1981) Quantitative tests for plate tectonics on Venus. *Geophysical Research Letters* 8: 1187–1190.
- Kaula WM, Schubert G, Lingenfelter RE, Sjogren WL, and Wollenhaupt WR (1974) Apollo laser altimetry and inferences as to the lunar structure. *Proceedings of the 5th Lunar and Planetary Science Conference*, Vol. 3, pp. 3049–3058. Pergamon Press.
- Kavner A, Duffy TS, and Shen G (2001) Phase stability and density of FeS at high pressures and temperatures: Implications for the interior structure of Mars. *Earth and Planetary Science Letters* 185: 25–33.
- Keil K (2002) Geological history of asteroid 4 Vesta. In: Bottke WF, Paolicchi P, Binzel RP, and Cellino A (eds.) *Asteroids III*, pp. 573–584. Tucson, AZ: University of Arizona Press.
- Khan A and Mosegaard K (2002) An inquiry into the lunar interior: A nonlinear inversion of the Apollo lunar seismic

- data. *Journal of Geophysical Research* 107(E6): 50036 (doi:10.1029/2001JE001658).
- Khan A, Mosegaard K, and Rasmussen KL (2000) A new seismic velocity model for the Moon from a Monte Carlo inversion of the Apollo lunar seismic data. *Geophysical Research Letters* 27: 1591–1594.
- Khan A, Mosegaard K, Williams JG, and Lognonné P (2004) Does the moon possess a molten core? Probing the deep lunar interior using results from LLR and Lunar Prospector. *Journal of Geophysical Research* 109(E09): E09007 (doi:10.1029/2004JE002294).
- Khan A, MacLennan J, Taylor SR, and Connolly JAD (2006a) Are the Earth and the moon compositionally alike? Inferences on lunar composition and implications for lunar origin and evolution from geophysical modeling. *Journal of Geophysical Research* 111(E05): E05005 (doi:10.1029/2005JE002608).
- Khan A, Connolly JAD, Olsen N, and Mosegaard K (2006b) Constraining the composition and thermal state of the moon from an inversion of electromagnetic lunar day-side transfer functions. *Earth and Planetary Science Letters* 248: 579–598.
- Khan A, Connolly JAD, MacLennan J, and Mosegaard K (2007) Joint inversion of seismic and gravity data for lunar composition and thermal state. *Geophysical Journal International* 168: 243–258.
- Khurana KK, Kivelson MG, Stevenson DJ, et al. (1998) Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395: 777–780.
- Kivelson MG, Bagenal F, Kurth WS, Neubauer FM, Paranicas C, and Saur J (2004) Magnetospheric interactions with satellites. In: Bagenal F, Dowling T, and McKinnon W (eds.) *Jupiter, The Planet, Satellites, and Magnetosphere*, pp. 513–536. Cambridge UK: Cambridge University Press.
- Kivelson MG, Khurana KK, Stevenson DJ, et al. (1999) Europa and Callisto: Induced or intrinsic fields in a periodically varying plasma environment. *Journal of Geophysical Research* 104(A3): 4609–4626.
- Kivelson MG, Khurana KK, Russell CT, Volwerk M, Walker RJ, and Zimmer C (2000) Galileo magnetometer measurements: A stronger case for a subsurface ocean at Europa. *Science* 289: 1340–1343.
- Kivelson MG, Khurana KK, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.
- Kleine T, Münker C, Mezger K, and Palme H (2002) Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf-W chronometry. *Nature* 418: 952–955.
- Knapmeyer M, Oberst J, Hauber E, Wählisch M, Deuchler C, and Wagner R (2006) Working models for spatial distribution and level of Mars's seismicity. *Journal of Geophysical Research* 111: E11006 (doi:10.1029/2006JE002708).
- Konopliv AS, Asmar SW, Carranza E, Sjogren WL, and Yuan DN (2001) Recent gravity models as a result of the Lunar Prospector mission. *Icarus* 150: 1–18.
- Konopliv AS, Banerdt WB, and Sjogren WL (1999) Venus gravity: 180th degree and order model. *Icarus* 139: 3–18.
- Konopliv AS, Binder AB, Hood LL, Kucinskas AB, Sjogren WL, and Williams JG (1998) Improved gravity field of the Moon from Lunar Prospector. *Science* 281: 1476–1480.
- Konopliv AS and Yoder CF (1996) Venusian k_2 tidal Love number from Magellan and PVO tracking data. *Geophysical Research Letters* 23: 1857–1860.
- Konopliv AS, Yoder CF, Standish EM, Yuan DN, and Sjogren WL (2006) A global solution for the Mars static and seasonal gravity, Mars orientation, Phobos and Deimos masses, and Mars ephemeris. *Icarus* 182: 23–50.
- Korotev RL (2005) Lunar geochemistry as told by lunar meteorites. *Chemie der Erde-Geochemistry* 65: 297–346.
- Kovach RL and Chyba CF (2001) Seismic detectability of a subsurface ocean on Europa. *Icarus* 150: 279–287.
- Kucinskas AB and Turcotte DL (1994) Isostatic compensation of equatorial highlands on Venus. *Icarus* 112: 104–116.
- Kuskov OL (1995) Constitution of the Moon: 3. Composition of the middle mantle from seismic data. *Physics of the Earth and Planetary Interiors* 90: 55–74.
- Kuskov OL (1997) Constitution of the Moon: 4. Composition from the mantle from seismic data. *Physics of the Earth and Planetary Interiors* 102: 239–257.
- Kuskov OL and Fabrichnaya OB (1995) Constitution of the Moon: 2. Composition and seismic properties of the lower mantle. *Physics of the Earth and Planetary Interiors* 83: 197–216.
- Kuskov OL and Galimzyanov RF (1986) Thermodynamics of stable mineral assemblages of the mantle transition zone. In: Saxena SK (ed.) *Chemistry and Physics of the Terrestrial Planets*, pp. 310–361. New York: Springer-Verlag.
- Kuskov OL and Kronrod VA (1998) Constitution of the Moon: 5. Constraints on composition, density, temperature, and radius of a core. *Physics of the Earth and Planetary Interiors* 107: 285–306.
- Lahiri BN and Price AT (1939) Electromagnetic induction in non-uniform conductors, and the determination of the conductivity of the earth from terrestrial magnetic variations. *Philosophical Transactions of the Royal Society of London, Series A* 237: 509–540.
- Langevin Y (1997) The regolith of Mercury: Present knowledge and implications for the Mercury Orbiter mission. *Planetary and Space Sciences* 45: 31–37.
- Lay T, Heinz D, Ishii M, et al. (2005) Multidisciplinary impact of the deep mantle phase transition in perovskite structure. *EOS Transactions of the American Geophysical Union* 86: 1–5.
- Lee S, Zanolin M, Thode AM, Pappalardo RT, and Makris NC (2003) Probing Europa's interior with natural sound sources. *Icarus* 165: 144–167.
- Lemoine FG, Smith DE, Rowlands DD, et al. (2001) An improved solution of the gravity field of Mars (GMM-2B) from Mars Global Surveyor. *Journal of Geophysical Research* 106: 23359–23376.
- Lewis JS (1988) Origin and composition of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 651–666. Tucson, AZ: University of Arizona Press.
- Liebermann RC (2005) The future of high-pressure mineral physics. *EOS Transactions of the American Geophysical Union* 86(40): 365–373.
- Lin RP, Mitchell DL, Curtis DW, et al. (1998) Lunar surface magnetic fields and their interaction with the solar wind: Results from Lunar Prospector. *Science* 281: 1480–1484.
- Lingenfelter RE and Schubert G (1973) Evidence for convection in planetary interiors from first-order topography. *Earth, Moon and Planets* 7: 172–180.
- Lognonné P (2005) Planetary seismology. *Annual Review of Earth and Planetary Sciences* 33: 571–604.
- Lognonné P and Mosser B (1993) Planetary seismology. *Survey in Geophysics* 14: 239–302.
- Lognonné P, Giardini D, and Banerdt B (2000) The NetLander very broad band seismometer. *Planetary and Space Sciences* 48: 1289–1302.
- Lognonné P, Gagnepain-Beyneix J, and Chenet H (2003) A new seismic model of the Moon: Implications for structure, thermal evolution and formation of the Moon. *Earth and Planetary Science Letters* 211: 27–44.
- Longhi J, Knittle E, Holloway JR, and Wänke H (1992) The bulk composition, mineralogy and internal structure of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 184–208. Tucson, AZ: University of Arizona Press.

- Margot JL, Campbell DB, Jurgens RF, and Slade MA (1999) Topography of the lunar poles from radar interferometry: A survey of cold trap locations. *Science* 284: 1658–1660.
- McDonough WF and Sun S-S (1995) Composition of the Earth. *Chemical Geology* 120: 223–253.
- McKenzie D and Nimmo F (1997) Elastic thickness estimates for Venus from line of sight accelerations. *Icarus* 130: 198–216.
- McKenzie D, Ford PG, Johnson C, et al. (1992) Features on Venus generated by plate boundary processes. *Journal of Geophysical Research* 97(E8): 13533–13544.
- McKinnon WB, Zahnle KJ, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Models and observations. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 969–1014. Tucson, AZ: University of Arizona Press.
- McSween HY (1994) What we have learned about Mars from SNC meteorites. *Meteoritics* 29: 757–779.
- Meissner RO and Vetter UR (1979) Relationship between the seismic quality factor Q and the effective viscosity η^* . *Journal of Geophysics* 45: 147–158.
- Mizutani H (1995) Lunar interior exploration by Japanese lunar penetrator mission, Lunar-A. *Physics of the Earth* 43: 657–670.
- Moore WB and Schubert G (1997) Venusian crustal and lithospheric properties from nonlinear regressions of highland geoid and topography. *Icarus* 128: 415–428.
- Moore WB, Schubert G, Anderson JD, and Spencer JR (2006) The interior of Io. In: Lopes RMC and Spencer JR (eds.) *Io after Galileo*, pp. 89–108. Chichester, UK: Springer Verlag.
- Muhleman DO, Grossman AW, and Butler BJ (1995) Radar investigation of Mars, Mercury and Titan. *Annual Review of Earth and Planetary Sciences* 23: 337–374.
- Murakami M, Hirose K, Kawamura K, Sata N, and Ohishi Y (2004) Post-perovskite phase transition in MgSiO₃. *Science* 304: 855–858.
- Nakamura Y (1983) Seismic velocity structure of the lunar mantle. *Journal of Geophysical Research* 88: 677–686.
- Nakamura Y (2005) Farside deep moonquakes and deep interior of the moon. *Journal of Geophysical Research* 110: E01001 (doi:10.1029/2004JE002332).
- Nakamura Y, Latham G, Lammlein D, Ewing M, Duennebier F, and Dorman J (1974) Deep lunar interior inferred from recent seismic data. *Geophysical Research Letters* 1: 137–140.
- Ness NF (1979) The magnetic fields of Mercury, Mars, and Moon. *Annual Review of Earth and Planetary Sciences* 7: 249–288.
- Neumann GA, Zuber MT, Smith DE, and Lemoine FG (1996) The lunar crust: Global structure and signature of major basins. *Journal of Geophysical Research* 101: 16841–16863.
- Neumann GA, Zuber MT, Wieczorek MA, McGovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research* 109: E08002 (doi:10.1029/2004JE002262).
- Nimmo F and McKenzie D (1998) Volcanism and tectonics on Venus. *Annual Review of Earth and Planetary Sciences* 26: 23–51.
- Nimmo F and Stevenson DJ (2000) Influence of early plate tectonics on the thermal evolution and magnetic field of Mars. *Journal of Geophysical Research* 105: 11969–11979.
- Nimmo F and Stevenson DJ (2001) Estimates of Martian crustal thickness from viscous relaxation of topography. *Journal of Geophysical Research* 106: 5085–5098.
- Nimmo F and Tanaka K (2005) Early crustal evolution of Mars. *Annual Review of Earth and Planetary Sciences* 33: 133–161.
- Nozette S, Rustan P, Pleasance LP, et al. (1994) The Clementine mission to the Moon: Scientific overview. *Science* 266: 1835–1839.
- Oganov AR and Ono S (2004) Theoretical and experimental evidence for a post-perovskite phase of MgSiO₃ perovskite in Earth's D' layer. *Nature* 430: 445–448.
- Olsen N (1999) Induction studies with satellite data. *Survey in Geophysics* 20: 309–340.
- Peale SJ (1976) Inferences from the dynamical history of Mercury's rotation. *Icarus* 28: 459–467.
- Peale SJ (1988) The rotational dynamics of Mercury and the state of its core. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 461–493. Tucson, AZ: University of Arizona Press.
- Phillips JL and Russell CT (1987) Upper limit on the intrinsic magnetic field of Venus. *Journal of Geophysical Research* 92(A3): 2253–2263.
- Phillips RJ (1994) Estimating lithospheric properties at Atla Regio, Venus. *Icarus* 112: 147–170.
- Pieters CM and Tompkins S (1999) Tsiolkovsky crater: A window into crustal processes on the lunar farside. *Journal of Geophysical Research* 104: 21935–21950.
- Pieters CM, Tompkins S, Head JW, and Hess PC (1997) Mineralogy of the mafic anomaly in the South Pole-Aitken basin: Implications for excavation of the lunar mantle. *Geophysical Research Letters* 24: 1903–1906.
- Platzman GW (1984) Planetary energy balance for tidal dissipation. *Reviews in Geophysics and Space Physics* 22: 73–84.
- Poirier JP and Liebermann RC (1984) On the activation volume for creep and its variation with depth in the Earth's lower mantle. *Physics of the Earth and Planetary Interiors* 35: 283–293.
- Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active South Pole of Enceladus. *Science* 311: 1393–1401.
- Press WH, Flannery BP, Teukolsky SA, and Vetterling WT (1992) *Numerical Recipes*. New York: Cambridge University Press.
- Ranalli G (1987) *Rheology of the Earth. Deformation and Flow Processes in Geophysics and Geodynamics*. Winchester, MA: Allen & Unwin.
- Rappaport NJ, Konopliv AS, Kucinskas AB, and Ford PG (1999) An improved 360 degree and order model of Venus topography. *Icarus* 139: 19–31.
- Righter K and Drake MJ (1996) Core formation in Earth's moon, Mars, and Vesta. *Icarus* 124: 513–529.
- Righter K and Drake MJ (1997) A magma ocean on Vesta: Core formation and petrogenesis of eucrites and diogenites. *Meteorites and Planetary Science* 32: 929–944.
- Rikitake T (1966) *Electromagnetism and the Earth's Interior*. Amsterdam, The Netherlands: Elsevier.
- Ringwood AE (1975) *Composition and Petrology of the Earth's Mantle*. New York: McGraw-Hill.
- Roberts JH and Zhong S (2006) Degree-1 convection in the Martian mantle and the origin of the hemispheric dichotomy. *Journal of Geophysical Research* 111: E06013 (doi:10.1029/2005JE002668).
- Roberts RG (1986) Global electromagnetic induction. *Survey in Geophysics* 8: 339–374.
- Russell CT (1980) Planetary magnetism. *Reviews of Geophysics* 18: 77–106.
- Russell CT, Coleman PJ, Jr., Fleming BK, et al. (1975) The fine-scale lunar magnetic field. *Proceedings of the 6th Lunar and Science Conference*, Vol. 3, pp. 2955–2969. The MIT Press.
- Russell CT, Coleman PJ, Jr., Lichtenstein BR, Schubert G, and Sharp LR (1973) Subsatellite measurements of the lunar magnetic field. *Proceedings of the 4th Lunar and Science Conference*, Vol. 3, pp. 2833–2845. The MIT Press.
- Sandwell DT and Schubert G (1992a) Evidence for retrograde lithospheric subduction on Venus. *Science* 257: 766–770.
- Sandwell DT and Schubert G (1992b) Flexural ridges, trenches, and outer rises around coronae on Venus. *Journal of Geophysical Research* 97(E10): 16069–16083.

- Sanloup C, Jambon A, and Gillet P (1999) A simple chondritic model of Mars. *Physics of the Earth and Planetary Interiors* 112: 43–54.
- Sanloup C, Guyot F, Gillet P, Fiquet G, Mezouar M, and Martinez I (2000) Density measurements of liquid Fe-S alloys at high-pressure. *Geophysical Research Letters* 27: 811–814.
- Schaber GG, Strom RG, Moore HJ, et al. (1992) Geology and distribution of impact craters on Venus – what are they telling us. *Journal of Geophysical Research* 97(E8): 13257–13301.
- Schubert G, Anderson JD, Spohn T, and McKinnon WB (2004) Interior composition, structure and dynamics of the Galilean satellites. In: Baggenstoss F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The planet, satellites and magnetosphere*, pp. 281–306. Cambridge, UK: Cambridge University Press.
- Schubert G, Anderson JD, Travis BJ, and Palguta J (in press) Enceladus: Present internal structure and differentiation by early and long term radiogenic heating. *Icarus*.
- Schubert G and Lingenfelter RE (1973) Martian centre of mass – Centre of figure offset. *Nature* 242: 251–252.
- Schubert G and Schwartz K (1969) A theory for the interpretation of lunar surface magnetometer data. *The Moon* 1: 106–117.
- Schubert G and Spohn T (1990) Thermal history of Mars and the sulfur content of its core. *Journal of Geophysical Research* 95: 14095–14104.
- Schubert G, Ross MN, Stevenson DJ, and Spohn T (1988) Mercury's thermal history and the generation of its magnetic field. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 429–460. Tucson, AZ: University of Arizona Press.
- Schubert G, Solomon SC, Turcotte DL, Drake MJ, and Sleep NH (1992) Origin and thermal evolution of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 147–183. Tucson, AZ: University of Arizona Press.
- Schubert G, Zhang K, Kivelson MG, and Anderson JD (1996) The magnetic field and internal structure of Ganymede. *Nature* 384: 544–545.
- Schubert G, Solomatov VS, Tackley PJ, and Turcotte DL (1997) Mantle convection and the thermal evolution of Venus. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1245–1287. Tucson, AZ: University of Arizona Press.
- Schubert G, Russell CT, and Moore WB (2000) Timing of the Martian dynamo. *Nature* 408: 666–667.
- Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge, UK: Cambridge University Press.
- Schuster A (1889) The diurnal variation of terrestrial magnetism. *Philosophical Transactions of the Royal Society of London Series A* 180: 467–518.
- Seidelmann PK, Abalkin VK, Bursa M, et al. (2002) Report of the IAU/IAG working group on cartographic coordinates and rotational elements of the planets and satellites: 2000. *Celestial Mechanics and Dynamical Astronomy* 82: 83–110.
- Seiff A (1983) Thermal structure of the atmosphere of Venus. In: Hunten DM, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 154–158. Tucson, AZ: University of Arizona Press.
- Sellers PC (1992) Seismic evidence for a low-velocity lunar core. *Journal of Geophysical Research* 97: 11663–11672.
- Senske DA, Schaber GG, and Stofan ER (1992) Regional topographic rises on Venus – Geology of Western Eistla Regio and comparison to Beta Regio and Atla Regio. *Journal of Geophysical Research* 97(E8): 13395–13420.
- Shampine LF and Gordon MK (1975) *Computer Solution of Ordinary Differential Equations. The Initial Value Problem*. San Francisco, CA: Freeman.
- Shearer CK, Hess PC, Wieczorek MA, et al. (2006) Thermal and magmatic evolution of the Moon. *Reviews in Mineralogy and Geochemistry* 60: 365–518.
- Siegfried RW and Solomon SC (1974) Mercury: Internal structure and thermal evolution. *Icarus* 23: 192–205.
- Simons M, Hager BH, and Solomon SC (1994) Global variations in the geoid/topography admittance of Venus. *Science* 264: 798–803.
- Sjogren WL, Banerdt WB, Chodas PW, et al. (1997) The Venus gravity field and other geodetic parameters. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1125–1162. Tucson, AZ: University of Arizona Press.
- Sloan ED (2003) Fundamental principles and applications of natural gas hydrates. *Nature* 426: 353–359.
- Smith DE and Zuber MT (1996) The shape of Mars and the topographic signature of the hemispheric dichotomy. *Science* 271: 184–187.
- Smith DE, Zuber MT, and Solomon SC (1999a) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1503.
- Smith DE, Sjogren WL, Tyler GL, Balmino G, Lemoine FG, and Konopliv AS (1999b) The gravity field of Mars: Results from Mars Global Surveyor. *Science* 286: 94–97.
- Smith JC and Born GH (1976) Secular acceleration of Phobos and Q of Mars. *Icarus* 27: 51–53.
- Smith JV, Anderson AT, Newton RC, et al. (1970) Petrologic history of the Moon inferred from petrography, mineralogy, and petrogenesis of Apollo 11 rocks. *Proceedings of the Apollo 11 Lunar and Science Conference*, Vol. 1, pp. 897–925. Pergamon Press.
- Smarekar SE (1994) Evidence for active hotspots on Venus from analysis of Magellan gravity data. *Icarus* 112: 2–26.
- Sohl F and Spohn T (1997) The interior structure of Mars: Implications from SNC meteorites. *Journal of Geophysical Research* 102: 1613–1635.
- Sohl F, Hussmann H, Schwentker B, Spohn T, and Lorenz RD (2003) Interior structure models and tidal Love numbers of Titan. *Journal of Geophysical Research* 108(E12): 5130 (doi:10.1029/2003JE002044).
- Sohl F, Schubert G, and Spohn T (2005) Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110: E12008 (doi:10.1029/2005JE002520).
- Solomon SC (2003) Mercury: The enigmatic innermost planet. *Earth and Planetary Science Letters* 216: 441–455.
- Solomon SC, Aharonson O, Aurnou JM, et al. (2005) New perspectives on ancient Mars. *Science* 307: 1214–1220.
- Sonett CP (1982) Electromagnetic induction in the moon. *Reviews in Geophysics and Space Physics* 20: 411–455.
- Sonett CP, Colburn DS, Dyal P, et al. (1971) Lunar electrical conductivity profile. *Nature* 230: 359–362.
- Sonett CP, Smith BF, Colburn DS, Schubert G, and Schwartz K (1972) The induced magnetic field of the Moon: Conductivity profiles and inferred temperatures. *Proceedings of the 3rd Lunar and Science Conference, Geochimica et Cosmochimica Acta, Supplement 3*, pp. 2309–2336. The MIT Press.
- Spencer JR, Pearl JC, Segura M, et al. (2006) Cassini encounters Enceladus: Background and the discovery of a south polar hot spot. *Science* 311: 1401–1405.
- Spohn T (1991) Mantle differentiation and thermal evolution of Mars, Mercury, and Venus. *Icarus* 90: 222–236.
- Spohn T, Sohl F, and Breuer D (1998) Mars. *Astronomical and Astrophysical Review* 8: 181–235.

- Spoohn T, Acuña MH, Breuer D, et al. (2001a) Geophysical constraints on the evolution of Mars. *Space Science Review* 96: 231–262.
- Spoohn T, Sohl F, Wieczkowski K, and Conzelmann V (2001b) The interior structure of Mercury: What we know, what we expect from BepiColombo. *Planetary and Space Sciences* 49: 1561–1570.
- Sprague AL, Kozlowski RWH, Witteborn FC, Cruikshank DP, and Wooden DH (1994) Mercury: Evidence for anorthosite and basalt from mid-infrared (7.3–13.5 μm) spectroscopy. *Icarus* 109: 156–167.
- Stacey FD (1977) Applications of thermodynamics to fundamental Earth physics. *Geophysical Survey* 3: 175–204.
- Stacey FD, Brennan BJ, and Irvine RD (1981) Finite strain theories and comparisons with seismological data. *Geophysical Survey* 4: 189–232.
- Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.
- Stofan ER, Hamilton VE, Janes DM, and Smrekar SE (1997) Coronae on Venus: Morphology and origin. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 931–965. Tucson, AZ: University of Arizona Press.
- Strom RG (1987) *Mercury: The Elusive Planet*. Washington, D.C.: Smithsonian Institute Press.
- Strom RG (1997) Mercury: An overview. *Advances in Space Research* 19: 1471–1485.
- Surkov YA (1983) Studies of Venus rocks by Veneras 8, 9, and 10. In: Hunten DM, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 154–158. Tucson, AZ: University of Arizona Press.
- Tarits P (1994) Electromagnetic studies of global geodynamic processes. *Survey in Geophysics* 15: 209–238.
- Tobie G, Grasset O, Lunine JI, Mocquet A, and Sotin C (2005) Titan's internal structure inferred from a coupled thermal-orbital model. *Icarus* 175: 496–502.
- Toksöz MN, Dainty AM, Solomon SC, and Anderson KR (1974) Structure of the Moon. *Reviews of Geophysics* 12: 539–567.
- Tsuchiya T, Tsuchiya J, Umemoto K, and Wentzcovitch RM (2004) Phase transition in MgSiO_3 perovskite in the earth's lower mantle. *Earth and Planetary Science Letters* 224: 241–248.
- Turcotte DL (1993) An episodic hypothesis for Venusian tectonics. *Journal of Geophysical Research* 98(E9): 17061–17068.
- Valencia D, O'Connell RJ, and Sasselov DD (2006) Internal structure of massive terrestrial planets. *Icarus* 181: 545–554.
- Valencia D, Sasselov DD, and O'Connell RJ (2007) Radius and structure of the first super-earth planet. *Astrophysical Journal* 656: 545–551.
- Van Hemelrijck E and Vercheval J (1981) Some aspects of the solar radiation incident at the top of the atmospheres of Mercury and Venus. *Icarus* 48: 167–179.
- Van Hoolst T and Jacobs C (2003) Mercury's tides and interior structure. *Journal of Geophysical Research* 108(E11): 5121 (doi:10.1029/2003JE002126).
- Van Hoolst T, Dehant V, Roosbeek F, and Lognonné P (2003) Tidally induced surface displacements, external potential variations, and gravity variations on Mars. *Icarus* 161: 281–296.
- van Thienen P, Rivoldini A, Van Hoolst T, and Lognonné P (2006) A top-down origin for martian mantle plumes. *Icarus* 185: 197–210.
- Vasavada AR, Paige DA, and Wood SE (1999) Near-surface temperatures on Mercury and the Moon and the stability of polar ice deposits. *Icarus* 141: 179–193.
- Verhoeven O, Rivoldini A, Vacher P, et al. (2005) Interior structure of terrestrial planets: Modelling Mars' mantle and its electromagnetic, geodetic and seismic properties. *Journal of Geophysical Research* 110: E04009 (doi:10.1029/2004JE002271).
- Vilas F (1985) Mercury: Absence of crystalline Fe^{2+} in the regolith. *Icarus* 64: 133–138.
- Voorhies CV, Sabaka TJ, and Purucker M (2002) On magnetic spectra of Earth and Mars. *Journal of Geophysical Research* 107(E6): 5034 (doi:10.1029/2001JE001534).
- Wänke H and Dreibus G (1988) Chemical composition and accretion history of terrestrial planets. *Philosophical Transactions of the Royal Society of London Series A* 325: 545–557.
- Warell J and Blewett DT (2004) Properties of the Hermean regolith. V. New optical reflectance spectra, comparison with lunar anorthosites, and mineralogical modelling. *Icarus* 168: 257–276.
- Warren PH (1985) The magma ocean concept and lunar evolution. *Annual Review of Earth and Planetary Sciences* 13: 201–240.
- Wasson JT (1988) The building stones of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 692–708. Tucson, AZ: University of Arizona Press.
- Watt JP, Davies GF, and O'Connell RJ (1976) The elastic properties of composite materials. *Reviews of Geophysics* 14: 541–563.
- Watters TR and McGovern PJ (2005) Hemispheres apart: The martian crustal dichotomy. *EOS Transactions of the American Geophysical Union* 86(5): 46–47.
- Weitz CM and Basilevsky AT (1993) Magellan observations of the Venera and Vega landing site regions. *Journal of Geophysical Research* 98(E9): 17069–17097.
- Wessel P and Smith WHF (1991) Free software helps map and display data. *EOS Transactions of the American Geophysical Union* 72: 441; 445–446.
- Wieczorek MA and Phillips RJ (1998) Potential anomalies on a sphere: Applications to the thickness of the lunar crust. *Journal of Geophysical Research* 103: 1715–1724.
- Wieczorek MA and Phillips RJ (1999) Lunar multiring basins and the cratering process. *Icarus* 139: 246–259.
- Wieczorek MA and Phillips RJ (2000) The Procellarum KREEP Terrane: Implications for mare volcanism and lunar evolution. *Journal of Geophysical Research* 105: 20417–20430.
- Wieczorek MA and Zuber MT (2001) The composition and origin of the lunar crust: Constraints from central peaks and crustal thickness modeling. *Geophysical Research Letters* 28: 4023–4026.
- Wieczorek MA and Zuber MT (2004) Thickness of the martian crust: Improved constraints from geoid-to-topography ratios. *Journal of Geophysical Research* 109: E01009 (doi:10.1029/2003JE002153).
- Wieczorek MA, Zuber MT, and Phillips RJ (2001) The role of magma buoyancy on the eruption of lunar basalts. *Earth and Planetary Science Letters* 185: 71–83.
- Wieczorek MA, Jolliff BL, Khan A, et al. (2006) The constitution and structure of the lunar interior. *Reviews in Mineralogy and Geochemistry* 60: 221–364.
- Wilhelms DE (1976) Mercurian volcanism questioned. *Icarus* 28: 551–558.
- Wilhelms DE and Squyres SW (1984) The Martian hemispheric dichotomy may be due to a giant impact. *Nature* 309: 138–140.
- Williams JG, Boggs DH, Yoder CF, Ratcliff JT, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106: 27933–27968.
- Wood BJ (1993) Carbon in the core. *Earth and Planetary Science Letters* 117: 593–607.

- Wood JA, Dickey JS, Marvin UB, and Powell BN (1970) Lunar anorthosites and a geophysical model of the Moon. *Proceedings of the Apollo 11 Lunar and Science Conference*, Vol. 1, pp. 965–988 Pergamon Press.
- Wood JA, Anderson DL, Buck WR, et al. (1981) Geophysical and cosmochemical constraints on properties of mantles of the terrestrial planets. In: Kaula WM (ed.) *Basaltic Volcanism on the Terrestrial Planets*, pp. 633–699. New York: Pergamon.
- Yoder CF (1981) The free librations of a dissipative Moon. *Philosophical Transactions of the Royal Society of London Series A* 303: 327–338.
- Yoder CF (1982) Tidal rigidity of Phobos. *Icarus* 49: 327–346.
- Yoder CF (1997) Venusian spin dynamics. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 1087–1124. Tucson, AZ: University of Arizona Press.
- Yoder CF and Standish EM (1997) Measurements of the precession and rotation from Viking lander range data. *Journal of Geophysical Research* 102: 4065–4080.
- Yoder CF, Konopliv AS, Yuan DN, Standish EM, and Folkner WM (2003) Fluid core size of Mars from detection of the solar tide. *Science* 300: 299–303.
- Zhong S and Zuber MT (2001) Degree-1 mantle convection and the crustal dichotomy of Mars. *Earth and Planetary Science Letters* 189: 75–84.
- Zimmer C, Khurana KK, and Kivelson MG (2000) Subsurface oceans on Europa and Callisto: Constraints from Galileo magnetometer observations. *Icarus* 147: 329–347.
- Zuber MT (2001) The crust and mantle of Mars. *Nature* 412: 220–227.
- Zuber MT, Smith DE, Lemoine FG, and Neumann GA (1994) The shape and internal structure of the Moon from the Clementine mission. *Science* 266: 1839–1843.
- Zuber MT, Smith DE, Solomon SC, et al. (1992) The Mars Observer laser altimeter investigation. *Journal of Geophysical Research* 97: 7781–7797.
- Zuber MT, Solomon SC, Phillips RJ, et al. (2000) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

10.03 Planetary Seismology

P. Lognonné, Institut de Physique du Globe de Paris, Saint Maur des Fossés, France

C. Johnson, University of British Columbia, Vancouver, BC, Canada

© 2007 Elsevier B.V. All rights reserved.

10.03.1	Introduction	69
10.03.2	Lunar Results	74
10.03.2.1	The Apollo PSE Data	74
10.03.2.2	Seismic Velocity Structure: Crust and Mantle	77
10.03.2.3	Very Deep Interior and Joint Seismic/Gravity Inversions	82
10.03.2.4	Mineralogical and Thermal Interpretation of Lunar Seismic Models	83
10.03.3	Seismic Activity of the Moon and Terrestrial Planets	86
10.03.3.1	Internal Seismic Activity	86
10.03.3.2	External Seismic Activity: Artificial and Natural Impacts	91
10.03.4	Atmospheric Seismology	93
10.03.4.1	Theoretical Background	93
10.03.4.2	Mars Hum and Martian Atmospheric Sources	98
10.03.4.3	Venus Atmospheric Seismology	100
10.03.4.4	Giant Planets Seismology	102
10.03.5	The Challenge: Mars Seismology	106
10.03.5.1	Interior Structure of Mars	106
10.03.5.2	Martian Seismic Noise	108
10.03.5.3	Body-Wave Detection	110
10.03.5.4	Normal Mode Excitation and Tidal Observations	111
10.03.5.5	Surface Waves	112
10.03.6	Concluding Remarks	112
	References	116

10.03.1 Introduction

One hundred and twenty years after the first detection of a remote seismic event by [Von Rebeur-Paschwitz \(1889\)](#) and 60 years after the elaboration of complete mean seismic models of the Earth from body waves traveltimes ([Bullen, 1947](#)), seismology is generally accepted as the best geophysical tool for the determination of the internal structure of a planet. Seismology has led to a revolution in Earth science, especially since the advent of three-dimensional (3-D) tomographic models of the Earth's mantle (e.g., [Dziewonski et al., 1977](#); [Woodhouse and Dziewonski, 1984](#)) that depict the shape of major discontinuities and reflect convection patterns and/or lateral variations in temperature or mineralogy in Earth's mantle. For a complete and extensive description of seismology applied to Earth, see Volume 1.

We are a long way from such a level of detail in planetary seismology and the seismic identification

of the Earth's core, realized by R. D. Oldham one century ago ([Oldham, 1906](#)), remains the only example of the detection of seismic waves refracted by a planetary core. The same is true for the detection of the normal modes of a planet, the first successful detection on Earth following the great Chilean earthquake in 1960 (e.g., [Bennioff et al., 1961](#)). As a consequence (see Chapter 10.02), we have today no precise measurement of the core radius of the Moon and Mars, no direct information on the Martian mantle structure, including on the discontinuities, no direct measurements on the Martian seismic activity, no direct measurement of the mean crustal thickness of Mars, etc.

The use of seismometers in planetary exploration was proposed early in the history of space missions (e.g., [Press et al., 1960](#)). Almost 45 years after the launch of the first seismometer to a telluric body (Ranger 3, in 1962), success in planetary and small-body seismology has been limited to Earth's Moon. The Apollo program deployed a network of

four seismic stations (Latham *et al.*, 1969, 1970a, 1970b, 1971) and the short-lived Apollo 11 seismometer for passive monitoring; three active seismic experiments were also conducted on Apollo 14–16 and 17 (Watkins and Kovach, 1972; Kovach and Watkins, 1973a). The 13 other extraterrestrial seismometer experiments that launched successfully never recorded any quakes (see Table 1 for a complete list). For the Moon, this includes the seismometers (Press *et al.* 1960; Lehner *et al.* 1962) onboard the three Ranger Lunar probes, that were lost with their missions in the early 1960s, the seismometer lost with the cancellation of the Moon landing of Apollo 13 and the gravimeter onboard the last Apollo 17 mission, which failed to operate and had been expected to extend the previous Apollo instruments with long-period seismic data. Several attempts have been made to conduct seismic experiments on Mars. Two seismometers were deployed by the Viking mission; however, one was never unlocked and the other provided no convincing event detection after 19 months of nearly continuous operation (Anderson *et al.*, 1977a, 1977b). Both the OPTIMISM seismometers (Lognonné *et al.*, 1998a) onboard the Small Surface Stations (Linkin *et al.*, 1998), and the Kamerton short-period seismometers (Kravroshkin and Tsyplakov, 1996) onboard the two penetrators were lost with the Mars 96 mission, after launch. Finally, for small bodies, the two short-period seismometers onboard the Phobos landers (Surkov, 1990) never reached Phobos. The SESAME/CASSE (Kochan *et al.*, 2000) acoustic experiment, onboard the Rosetta lander en route toward comet 67P/Churyumov-Gerasimenko, is therefore today the only experiment which might gather some seismic signals on another solid body other than Earth.

The seismometers developed for early missions used pioneering techniques as compared with then-state-of-the-art Earth instruments. The Ranger seismometer, proposed by Press *et al.* (1960) and designed by the California Institute of Technology (Lehner *et al.*, 1962) for the first missions to land on the Moon, was designed to survive up to a 3000 g shock ($1g = 9.81 \text{ m s}^{-2}$). This was a vertical-axis seismometer with a free frequency of 1 Hz and a mass of 3 kg, and was one of the first digital instruments. Although the three Ranger probes failed, the technology was reused for operating one of the first terrestrial digital seismometers used at Caltech (Miller, 1963). A new experiment was proposed for the *Surveyor* mission, including a three-axis long-period seismometer and a short-period vertical-axis

seismometer, built at the Lamont Doherty Geological Observatory (Sutton and Latham, 1964), with a total mass of 11.5 kg. This was later descoped to a single short-period vertical-axis seismometer (Sutton and Steinbacher, 1967), and then canceled in the Surveyor program. Although this effort had to await the Apollo missions for flight to the Moon, it was reused in terrestrial seismology, notably in the first ocean-bottom seismometers deployed in the 1970s. The deployment of seismometers was also planned for the Soviet Lunakhod, using a 1 Hz vertical fused quartz seismometer. This experiment was cancelled (Osika and Daragan, personal communication, 1990).

The installation of the first operational seismometer on a planetary body other than the Earth was achieved in Jul 1969, during the Apollo 11 mission. The Passive Seismic Experiment (PSE), consisted of a tri-axis long period (LP) seismometer, with a resonance period of 15 s, and one vertical short period (SP) seismometer, with a resonance period of 1 s. The total mass of both instruments plus the electronic and thermal control module was 11.5 kg and the consumption power was between 4.3 and 7.4 W (Latham *et al.*, 1969, 1970a, 1970b). These seismometers were extremely sensitive, capable of detecting a displacement of $3 \times 10^{-10} \text{ m}$ at frequencies of 0.1–1 Hz, for the LP in flat mode; $0.5 \times 10^{-10} \text{ m}$ at 0.45 Hz for the LP in peaked mode; and $0.5 \times 10^{-10} \text{ m}$ for the SP seismometer at 8 Hz. The corresponding nominal acceleration response curves are shown in Figure 1, as well as the noise recorded on the Moon. Practically, these instruments were unable to detect the continuous microseismic noise of the Moon, and detected only a significant background noise around the sunrise and sunset, associated with thermal effects located in the shallow subsurface. The Moon noise recorded on the vertical component seismometer is as low as $10^{-10} \text{ ms}^{-2} \text{ Hz}^{-1/2}$ at frequencies in the range 0.1–1 Hz (Earth's microseismic peak). This allowed the detection of about 12 500 seismic signals between Jul 1969 and Sep 1977 (a mean of four quakes per day); many more unreported events were seen on the SP instruments. The most recent LP event catalog is available online at the University of Texas (Nakamura *et al.*, 2004).

The first PSE experiment was operated by solar panels, and survived the first lunar night but failed, due to high temperatures, during the second lunar day (Latham *et al.*, 1969, 1970a, 1970b). Subsequent seismometers used a radio thermal generator

Table 1 Summary and history of planetary seismology experiments

Mission	Launch	Major mission events	Instrument description	Seismometer deployment	Reference
Ranger 3	1962-01-26	Failure due to the booster. Moon missed	Vertical axis seismometer, with a free frequency of 1 Hz. (Mass: 3.36 kg)	Seismometer in a lunar capsule designed for a $130\text{--}160 \text{ km h}^{-1}$ landing. Batteries powered for 30 days of operations	Lehner et al. (1962)
Ranger 4	1962-04-23	Failure of spacecraft central processor. Moon crash.			
Ranger 5	1962-10-18	Failure in the spacecraft power system. Moon missed.			
Surveyor	1966-1968	The seismometer was finally deselected from the payload of the Surveyor missions	Single short-period vertical axis seismometer (mass: 3.8 kg, power: 0.75 W)	Fixed to the lander.	Sutton and Steinbacher (1967).
Apollo 11	1969-7-16	Successful installation. Powered by solar panel, worked during the first lunation and stopped after 21 days	Passive seismic experiment (PSE). triaxis Long-Period seismometer (LP) and one vertical Short-Period (SP) seismometer, with resonance periods of 15 s and 1 s, respectively. (mass: 11.5 kg, power: 4.3–7.4 W)	Installation performed by crew. Seismometers were manually leveled and oriented with bubble level and sun compass. A sun protection/thermal shroud was covering the instruments. Power was delivered by a Plutonium thermal generator for A12-14-15-16	Latham et al. (1969, 1970a, 1970b)
Apollo 12	1969-11-14	Successful installation of a network of 4 stations. For all but the Apollo 12 SP seismometer and Apollo 14 vertical LP seismometer operated until the end of Sep 1977, when all were turned off after command from the Earth. 26.18 active station years of data collected.			
Apollo 14	1971-01-31				
Apollo 15	1971-07-26				
Apollo 16	1972-04-16				
Apollo 13	1970-4-11	Moon landing aborted. No installation of the PSE experiment but lunar crash of the Apollo 13 Saturn-IV upper stage recorded by the A12 PSE.			
Apollo 14	1971-01-31	Successful installation and operation of the active seismic experiments. Seismic sources were thumper devices containing 21 small explosive sources and a rocket grenade launcher with four sources exploding up to 1500 m on A-14 and A-16. Eight sources were used containing up to 2722 g of explosive and deployed at 3500 m by astronauts	String of three geophones on A-14 and A16 and on four geophones on A-17. Frequency was 3–250 Hz.	Geophones were anchored into the surface by short spikes as they were unreeled from the thumper/geophone assembly.	Watkins and Kovach (1972)
Apollo 16	1972-04-16				
Apollo 17	1972-12-07				Kovach and Watkins (1973a)

(Continued)

Table 1 (Continued)

Mission	Launch	Major mission events	Instrument description	Seismometer deployment	Reference
Apollo 17	1972-12-07	Deployment of the Lunar Surface Gravimeter. The gravimeter was unable to operate properly due to an error in the design of the proof mass.	Gravimeter designed for gravity waves detection. Additional long-period vertical seismic output (10^{-11} lunar g resolution) for free oscillation detection, with a 16 Hz sampling.	Installation performed by crew.	Weber (1971)
<i>Viking Lander 1</i>	1975-08-20	Successful landing but instrument failure.	Short-period instrument, with an undamped natural period of 0.25 s, a mass of 2.2 kg, a size of $12 \times 15 \times 12$ cm and a nominal power consumption of 3.5 W.	The seismometer was installed on the Lander platform. No recentering was necessary since the three-axis seismometer had been designed to function even when tilted to up to 23°	Anderson et al. (1977a, 1977b)
<i>Viking Lander 2</i>	1975-09-09	Successful landing and 19 months of nearly continuous operation. Too high wind sensitivity associated to the elastic recovery of the <i>Viking</i> landing legs to the loading of the station by pressure fluctuations induced by the wind.			
Phobos 1-2	1988-07-07 1988-07-12	Respectively: Lost during transfer to Mars and Phobos; contact lost just before the final phase of lander deployment, after Mars orbit insertion		Instrument onboard the long-service lander.	Surkov (1990)
Mars 96-Small surface stations	1996-11-16	Failure of the Block-D propulsion system in parking orbit. Earth re-entry. Two small stations and two penetrators lost.	Long-period vertical axis seismometer (0.1–4 Hz, 0.405 kg for the sensor) combined to a magnetometer. 55 mW of power	Seismometer in the small surface station. Semi-hard landing (200g–20 ms). Nominal operations of one Martian year with 90th first days of nearly continuous mode with internal batteries	Lognonné et al. (1998a)
<i>Mars 96 Penetrators</i>			High-frequency seismometer (10–100 Hz, 0.3 kg, 20 mW)	Seismometer in the penetrator. Hard landing. Nominal operations of one Martian year.	Kravoshkin and Tsyplakov (1996)
Rosetta	2004-03-04	Landing on the comet 67P/Churyumov-Gerasimenko planned a few months after rendezvous, expected on 22-05-2014	CASSE/SESAME experiment: High-frequency accelerometer covering the frequency bandwidth \sim 10–20 kHz.	Instrument mounted on the lander	(Kochan et al. 2000)

Successful experiments, leading to detection and interpretation of data in terms of quakes, are indicated by their name in bold, in the Mission column. Bold and italic names are for successful deployment, but without clear event detection or interpretation. Italic names are for unsuccessful missions. For information on instruments for future missions see Lognonné (2005).

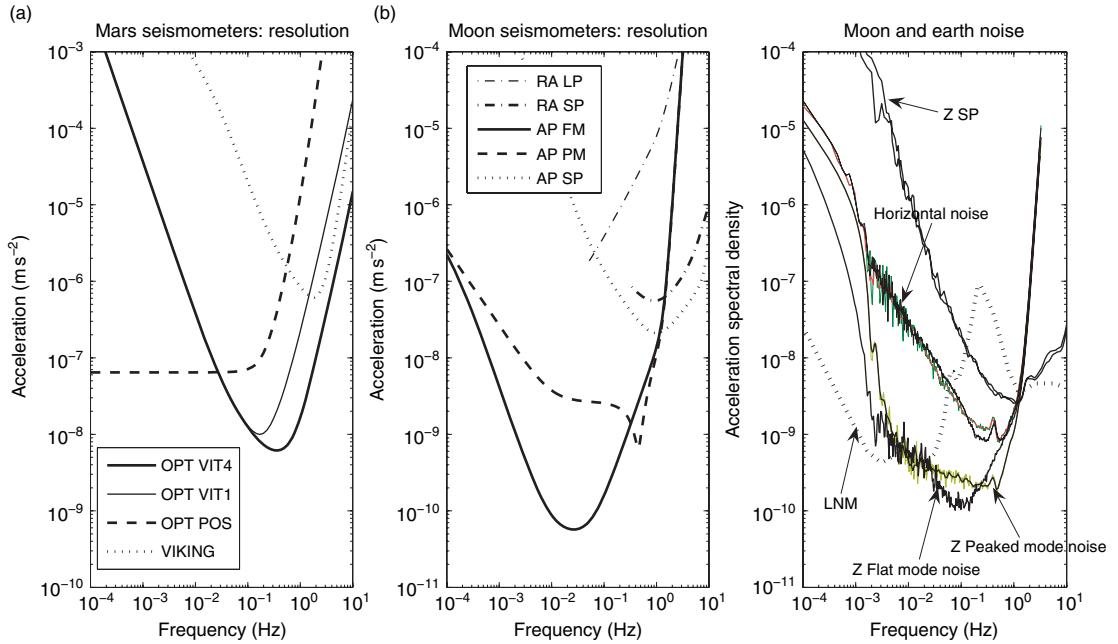


Figure 1 Response curves of the past (a, left) Mars and (b, middle) Moon seismometers. (a) shows the resolution of the Viking seismometer, the Optimism seismometer, in the long-period position mode (POS), in the velocity mode OPT VIT1 at 1 sample per second (sps), and at 4 sps (OPT VIT4). (b) shows the Moon case, with the long (RA LP) and short (RA SP)-period analog outputs of the Ranger seismometer, the Flat (AP FM) and peaked (AP PM) mode of the Apollo LP seismometer and the short-period seismometer (AP SP). The noise levels, recorded on the Moon by the different channels of the Apollo seismometers are compared with the earth-based Low Noise model of Peterson (1993). Noise levels are from the Apollo 12 instrument for the long-period (LP) component and from the Apollo 14 instrument for the short-period (SP) component. These noise levels are likely an upper estimate of the Moon noise and may be related to the instruments.

allowing continuous day and night operation. A seismic network of four stations (Figure 2) was installed by Apollo 12, 14, 15, 16. This network has been of critical importance in understanding the internal structure of the Moon, as has been discussed in several other lunar and seismological reviews (e.g., Hood, 1986; Hood and Zuber, 2000; Lognonné, 2005). Due to its low seismic noise, the Moon has been also proposed as a right place for the detection of astrophysical gravitational waves (e.g., Gusev and Kurlachev, 1976; Tobias, 1978) and a gravimeter designed with that respect was onboard Apollo 17. It was, however, never completely deployed and failed to operate. More recently, the possibility for strange quark matter detection (Banerdt *et al.*, 2006) has been proposed, reinitiating the possible interest for using on the Moon seismic or gravity sensors for astrophysics purposes.

In this chapter, we summarize the experiment and seismological data in Section 10.03.2. We devote some attention to 1-D seismic velocity models for the Moon's crust and mantle, and discuss differences among current models, especially their likely causes

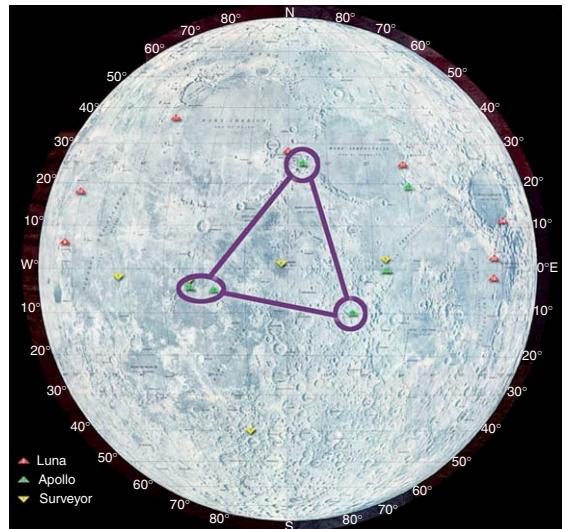


Figure 2 Configuration of the Apollo Seismic Network, A12 is in the Oceanus Procellarum area at 3.01°S , 23.42°W , A14 is located near the crater Fra Mauro at 3.64°S , 17.47°W , A15 is at the foot of the Apennine mountain range at 26.13°N , 3.63°E and A16 is just north of the crater Dolland at 8.97°S , 15.50°E . Note that all stations were on the near side, rendering core seismic studies almost impossible. Base map credit NASA.

and implications. We explore what has been learned about the mineralogical structure of the Moon, as a consequence of the integration of the Apollo seismic data with other geophysical and petrological constraints. In Section 10.03.3 we discuss the seismic activity of the Moon in the context of other terrestrial planets.

We address the field of atmospheric seismology in Section 10.03.4. The basic theory for coupling of solid body modes to the atmosphere and ionosphere is presented. We discuss the giant planets, especially Jupiter, for which the atmospheric signals associated with continuously excited global oscillations have been monitored from Earth for about 20 years (Lognonné and Mosser, 1993; Mosser, 1995). We develop a comparative study of the atmosphere/interior seismic coupling for Mars and Venus and discuss the prospect of remotely sensed seismology.

The seismic exploration of Mars has been much less successful than that of the Moon, starting with the deployment in 1976 of the Viking seismometers (Anderson *et al.*, 1977a, 1977b). Only the seismometer on board Viking 2 worked, the seismometer on Viking 1 lander failed to unlock. The sensitivity of the Viking 2 seismometer was one order of magnitude less than the SP Apollo seismometer for periods shorter than 1 s, and five orders less than the LP seismometer, for periods longer than 10 s (see Figure 1). No convincing event detection was made during the 19 months of nearly continuous operation of the Viking Lander 2 seismometer. As shown by Goins and Lazarewicz (1979), this was probably related to the inadequate sensitivity of the seismometer in the frequency bandwidth of teleseismic body waves, as well as high sensitivity to wind noise (Nakamura and Anderson, 1979). In 1996, the Mars 96 mission was launched. Each of the two small stations (Linkin *et al.*, 1998) carried the OPTIMISM seismometer in their payload (Lognonné *et al.*, 1998a). The sensitivity of the OPTIMISM seismometer was improved by about two orders of magnitude relative to the Viking seismometers at frequencies of 0.5 Hz, and thus better adapted to teleseismic body-wave detection. The Mars 96 mission was, however, lost shortly after its launch and the seismic exploration of Mars remains an outstanding avenue of investigation in solar system science. We return to the discussion of Mars seismology in Section 10.03.5, as it leads the way in potential future surface-based exploration of other bodies. We discuss current understanding of the Martian interior along with noise levels and other issues relevant to marsquake detection.

10.03.2 Lunar Results

10.03.2.1 The Apollo PSE Data

The lunar PSE data, recorded by the Apollo Lunar Surface Experiment Package (ALSEP) from 1969 to 1977, provide a unique and valuable resource for constraining the interior structure of the Moon. The Apollo seismic ‘network’ comprised four stations at Apollo sites 12, 14, 15, and 16. Stations 12 and 14 were about 180 km apart, and formed one corner of an approximately equilateral triangle, with stations 15 and 16 at the other corners, each about 1100 km away (Figure 2). The lunar event catalog (Nakamura *et al.*, 1981) documents 12 558 events. Recorded events exhibit different signal characteristics, from four types of sources, which were originally classified as artificial impacts (9), meteoroid impacts (\sim 1700), shallow moonquakes (28), deep moonquakes (\sim 1360), and unclassified events. Many of the originally unclassified events have since been classified as deep moonquakes, with a total of 7245 deep moonquakes proposed by Nakamura (2003).

Compared with terrestrial quakes, moonquakes are small seismic moment events and generated small signals, most of the time with amplitudes smaller than the Earth’s microseismic noise, even in the best seismic vaults on Earth. Event identification was originally made by eye, using hard-copy printouts and overlays of the seismograms recorded at Apollo station 12, 14, 15, and 16. Criteria used to identify and classify seismic sources were the time interval from the start of a signal to its maximum amplitude (rise time), the dominance of low-frequency versus high-frequency content in the records, and the presence of compressional (P) and/or shear (S) waves (see review in Lammlein (1977) and details in Lammlein (1973) and Lammlein *et al.* (1974)). Shallow events are also referred to in the literature as high-frequency teleseismic (HFT) events, due to their unusually high-frequency content and poor depth location (Nakamura *et al.*, 1974, 1980; Nakamura, 1977). Analyses of data early during the ALSEP suggested that deep moonquakes have two important characteristics that were subsequently used to help identify additional such events: similar waveforms, and the presence of tidal periodicities in moonquake occurrence times (Ewing *et al.*, 1971; Lammlein, 1977). The main properties of the seismic sources, in term of seismic moment, stress drop, rupture times will be discussed in Section 10.03.3. Examples of

waveforms recorded at the Apollo stations from each of the four sources are shown in **Figure 3**. Events are, in general, apparent only after filtering of the seismic data; the raw waveform data are dominated

by large-amplitude signals at each station due to expansion and contraction of the thermal Mylar shroud at sunrise and sunset (Duennebier and Sutton, 1974, see **Figure 3** of Bulow *et al.*, 2005).

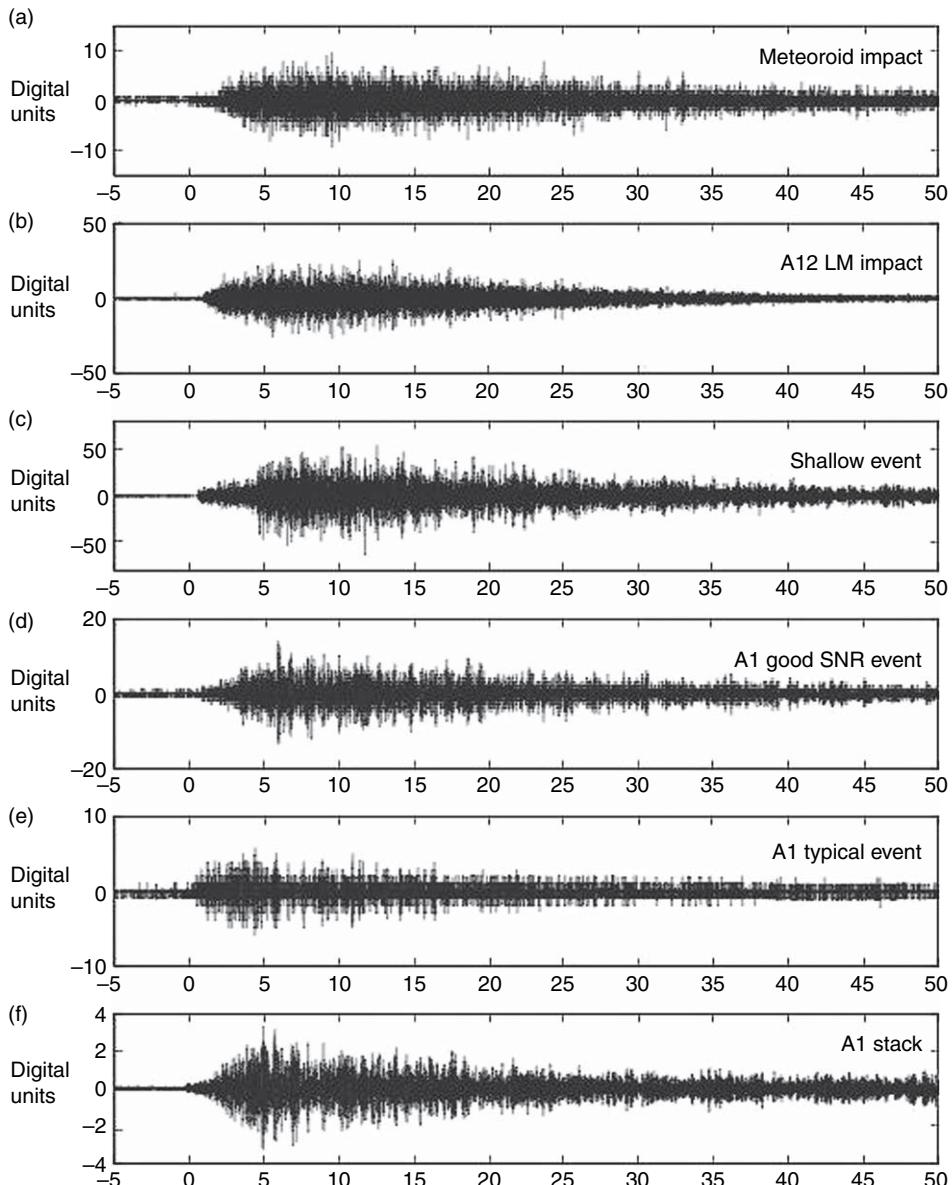


Figure 3 From top to bottom, examples of waveforms for natural and artificial impacts ((a) and (b)), shallow event (c), and for the three bottom ones ((d)–(f)) from deep moonquakes recorded on long-period channels by the Apollo Passive Seismic Network. All records span 55 min – x-axis time is in minutes, y-axis is in digital units. The following characteristics aided the identification of different seismic sources in the Apollo data: (1) the rise time (defined as time for event onset to maximum signal amplitude) of moonquakes is typically less than that of meteoroids – see the approximately 5 min rise time in (c), (d), and (e) above vs the more emergent arrival (perhaps 10 min rise time) in (a); (2) the large amplitude of the artificial impacts compared with natural events – see the scale for (b) vs (d) and (e) above. (Note that in our example we have chosen one of the largest-magnitude shallow events, and this does have comparable amplitude to the chosen artificial impact.) (3) Low amplitude and poor data quality of typical deep moonquakes – see (e). However, waveform repeatability of deep moonquakes from distinct source regions allows stacking to improve the signal-to-noise ratio (f).

During the Apollo era, digital data were extracted from the complete continuous PSE records in time windows encompassing identified events, and were stored as the ‘event’ waveform data set. This derived data set and the original continuous records are now available, with documentation, from the IRIS Data Management Center. Recent work (Nakamura, 2003) has classified a significant fraction of identified, but unclassified, events (Nakamura *et al.*, 1981) and has led to almost a factor of 5 increase in the number of catalogued deep moonquakes. Additionally, work by Bulow *et al.* (2005) shows that the waveform repeatability of deep moonquakes can be exploited to search for additional, previously unidentified events, associated with known deep source regions. Implementation of this algorithm has led to an approximately 35% increase in the number of events at the nine most seismically active (at the detection level of the Apollo seismometers) source regions (Bulow *et al.*, 2005, 2006, 2007). Despite the seemingly large number of lunar seismic events, the poor signal-to-noise ratio for emergent arrivals on most of the records means that it is possible to measure arrival times and accurately locate events only for a subset of data in the lunar catalogue (Figure 4). Small event magnitudes and restricted receiver locations mean that events for which traveltimes can be clearly identified mainly occur on the near side.

The Apollo passive seismic data set has contributed enormously to our understanding of the internal

structure of the Moon. The temporal and spatial distributions of lunar seismicity provide constraints on subsurface structure. Of particular importance is the evidence for brittle failure at depths of 700–1200 km, and likely concentrated in the 800–1000 km depth interval. Unfortunately, the small magnitude of these events, combined with the limited number and geographical distribution of receivers, means that it has not been possible to unambiguously establish a focal mechanism for any of the deep moonquakes, although this has been attempted (Koyama and Nakamura, 1980). Seismic velocity models for the crust and mantle (see Section 10.03.2.2) constrain the physical properties of the lunar interior (including temperature) and provide indirect constraints on mineralogy and the thermal evolution of the Moon. Studies of seismic attenuation indicate a dry lunar mantle (Nakamura and Koyama, 1982). Scattering of seismic energy, manifested as long coda in the observed waveforms, has been used to investigate shallow (upper crust and regolith) fracturing (e.g., Nakamura, 1977). Attempts have been made to investigate lunar normal modes (Khan and Mosegaard, 2001) to constrain the core size and state. However, these have been largely unsuccessful because of the limited bandwidth of the Apollo seismometers (Figure 1). In the following section, we discuss attempts to establish seismic velocity profiles for the lunar crust and mantle. Because of their potential to provide significant insights concerning

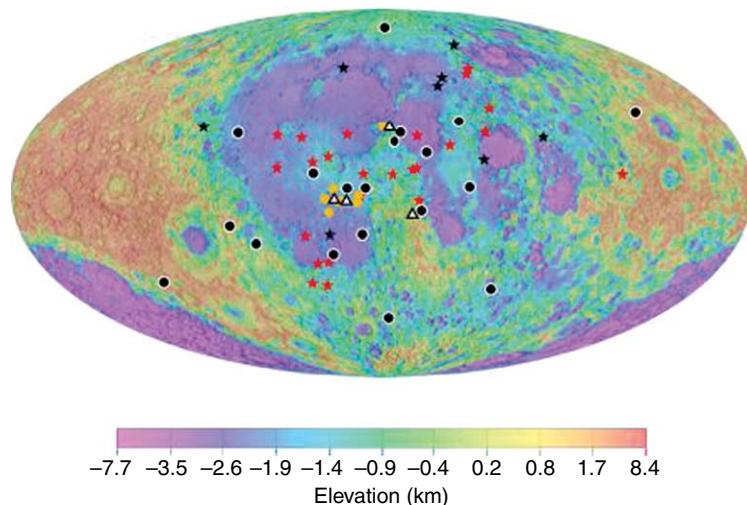


Figure 4 Seismic events located by Lognonné *et al.* (2003) and used in recent velocity models (Lognonné *et al.*, 2003; Gagnepain-Beyneix *et al.*, 2006), Apollo stations 12, 14, 15, and 16 are shown (white triangles), along with the locations of 24 deep event clusters (red stars), 8 shallow events (black stars), 19 meteoroid impacts (black circles), and 8 artificial impacts (orange circles). Background color map is Clementine topography (Aitoff equal area projection), shaded using a global digital image map (USGS map I-2769, 2002).

the lunar interior, these seismic velocity models have been used for many seismological, mineralogical, and thermal studies of the Moon.

10.03.2.2 Seismic Velocity Structure: Crust and Mantle

Due to scattering in the lunar seismograms, as well as the limited computational resources of the 1970s, most investigations of the lunar interior have used arrival times of body-wave phases to investigate plausible 1-D models for seismic velocity structure. (see Chapter 10.02 for a discussion of the lunar interior structure) This approach has been used with great success in terrestrial seismology, in a series of efforts leading to the longstanding Preliminary Reference Earth Model (PREM) for Earth's interior structure (Dziewonski and Anderson, 1981). Lunar studies use arrival times of direct P and S waves, picked from seismograms for as many channels and stations as possible. Individual seismograms are used for impacts and shallow moonquakes. For deep moonquake sources, all records of a specific source region are

stacked to improve the signal-to-noise ratio (Nakamura, 1983; Lognonné *et al.*, 2003). Source-to-receiver P and S wave traveltimes depend on receiver location (known), source location (generally unknown, though natural impact occur on the surface, and artificial impact locations are known), source time (unknown, except for artificial impacts), and interior velocity structure (unknown). The standard seismological approach is to simultaneously solve for source locations and times, along with interior velocity structure (e.g., Shearer, 1999).

P- and S-wave arrival times measured from the lunar seismograms have been either inverted or forward-modeled to estimate lunar seismic velocity profiles (e.g., Nakamura *et al.*, 1976; Goins *et al.*, 1981b; Nakamura, 1983; Khan *et al.*, 2000; Khan and Mosegaard, 2002; Lognonné *et al.*, 2003, Gagnepain-Beyneix *et al.*, 2006). As the construction of lunar seismic velocity models has been reviewed elsewhere (Lognonné, 2005), we do not repeat this overview here. Instead we restrict our discussion to four 1-D models for P- and S-wave velocities that illuminate the main unresolved questions concerning lunar interior structure (Figure 5).

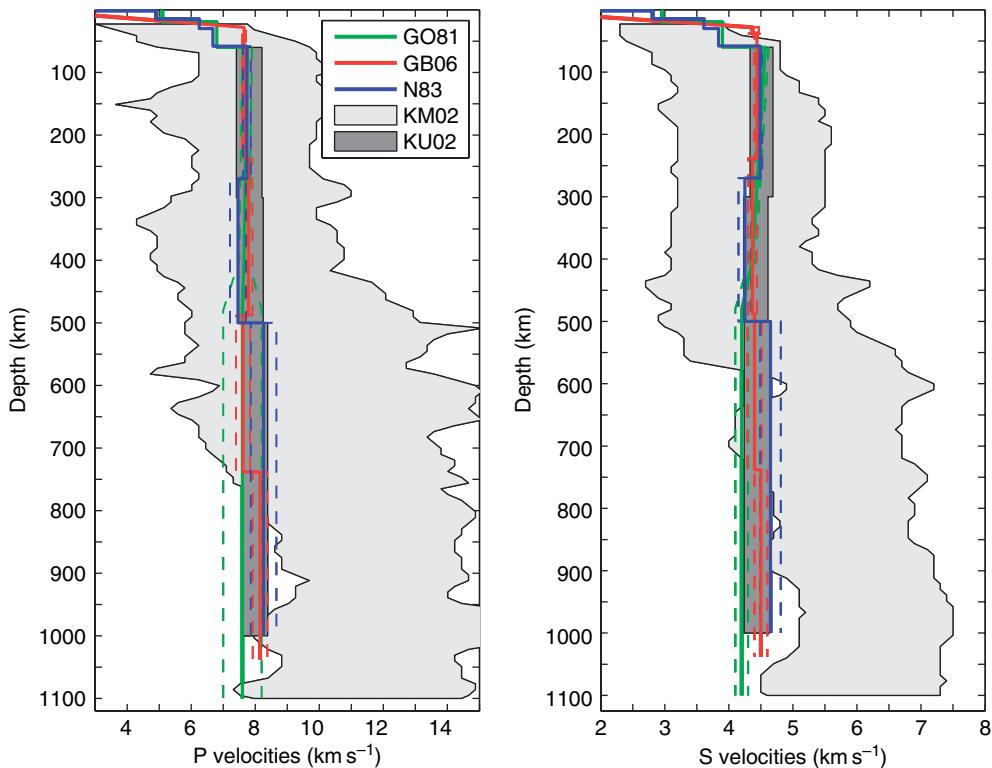


Figure 5 P- and S-wave velocity models vs depth: green, Goins *et al.* (1981b), red, Ganepain-Beyneix *et al.* (2006), blue, Nakamura (1983). The light gray zone is the 80% probability zone of Khan and Mosegaard (2002); The dark gray is the petrology/geochemistry model of Kuskov *et al.* (2002). Values for mantle velocities and associated 1σ uncertainties for each model can be found in table 1 of Lognonné (2005) and are plotted as colored dashed lines with the same colors for the three first models.

Figure 5 shows the P- and S-velocity models of Goins *et al.* (1981b), Nakamura (1983), Khan and Mosegaard (2002), (see also Khan *et al.* (2000) and Gagnepain-Beyneix *et al.* (2006)). We choose these studies since they exemplify state-of-the-art calculations immediately following the publication of the complete lunar event catalog (Goins *et al.*, 1981b; Nakamura, 1983), alternative modern probabilistic techniques applied to the early traveltimes data set (Khan *et al.*, 2000), and recent revised traveltimes picks and derived velocity models (Lognonné *et al.*, 2003; Gagnepain-Beyneix *et al.*, 2006). The models of Nakamura (1983), Khan *et al.* (2000), and Khan and Mosegaard (2002) have been used in mineralogical modeling of the Moon (see later), and Gagnepain-Beyneix *et al.* (2006) specifically discuss the thermal and mineralogical implications of their model.

The model of Nakamura (1983) (hereafter referred to as N83) uses traveltimes from a total of 81 events. A layered model was specified with four crustal layers and mantle interfaces at 270 and 500 km depth. The model minimized the root-mean-square misfit of the predicted and observed traveltimes; linearizing about the starting model. Nakamura (1983) stressed that the choice of interface depths and the specification of constant velocity layers was arbitrary. Khan *et al.* (2000) and Khan and Mosegaard (2002) used a Monte Carlo simulation approach to find, with a Bayesian probability view, families of models compatible with the Nakamura's (1983) traveltimes data set. We refer to Khan *et al.* (2000) and Khan and Mosegaard (2002) hereafter as KM02 collectively, as the approach and broad conclusions of these two studies are in agreement. Models were computed by perturbing an initial 56-shell model in which seismic velocity varied monotonically within each shell. Bayesian probabilistic tests were conducted to distinguish between two different depth ranges of the crust–mantle boundary. P- and S-wave traveltimes were repicked for 59 events by Lognonné *et al.* (2003), and used in models by Lognonné *et al.* (2003), and Gagnepain-Beyneix *et al.* (2006) (hereafter L03 and GB06, respectively) that specify 3–4 mantle layers and 4–6 crustal layers.

Crustal structure is primarily controlled by impacts. Seismic ray theory predicts that near-receiver impacts (mostly the artificial impacts) will have rays that turn within the crust. In addition, distant impacts produce seismic energy that travels through the crust near the source location, and again near the receiver location. Thus, in theory,

traveltimes of P and S waves should constrain crustal thickness and velocity structure within the crust, in particular near the longest-lived Apollo stations (12 and 14). Additional data that can be used to investigate crustal structure come from observations of S-to-P converted phases (Vinnik *et al.*, 2001); these have been studied at stations A12 and A16 (at A14 and A15, data were either missing or too noisy). Converted phases were detected only at the Apollo 12 station and originate at a seismic velocity discontinuity interpreted as the crust–mantle boundary. These types of studies are referred to as receiver function studies in the terrestrial seismological literature. The combination of differential traveltimes and relative amplitude of the primary and converted phases provide information on the depth to, and the impedance contrast across, the discontinuity. However, the lunar data set is less than ideal. Recent studies (L03, GB06) use 103 traveltimes from seven artificial and 19 natural impacts. The latter require estimates of three (latitude, longitude, and source time) source parameters, reducing the number of degrees of freedom to 46. Traveltimes and relative amplitudes for converted phases from only one receiver location are sufficiently reliable for incorporation into velocity profile modeling (L03, GB06).

Figure 5 shows significant variability among crustal velocity models, in particular in estimates of crustal thickness. By analogy with Earth, the lunar seismological crust–mantle interface is defined by a velocity discontinuity or steep gradient ($>0.1 \text{ km s}^{-1} \text{ km}^{-1}$), below which the P-wave velocity should attain a value of at least 7.6 km s^{-1} . Specifically, crustal thickness estimates are 58 km (N83, and see earlier work by Toksöz, 1974), 45 km (KM02), and 30 km (GB06) (**Figure 5**). The model of GB06 requires a 10 km brecciated zone beneath the crust. Velocity structure at depths of 30–60 km is quite sensitive to a few traveltimes data (see discussion in GB06). **Figure 6** shows sampling of the lunar interior by ray paths for the event and receiver locations in **Figure 4**. Ray path coverage is uneven with depth. Coverage of very shallow depths ($<60 \text{ km}$) is hard to see in **Figure 6**, but is poor. While uncertainties in crustal thickness remain, four main conclusions can be drawn from the seismological data. First, although the profiles shown in **Figure 5** are 1-D velocity profiles, the source–receiver geometries and seismic ray theory indicate that the averaged crustal thickness reflects primarily crustal structure at the Apollo 12 and 14 sites. Second, several recent tests indicate

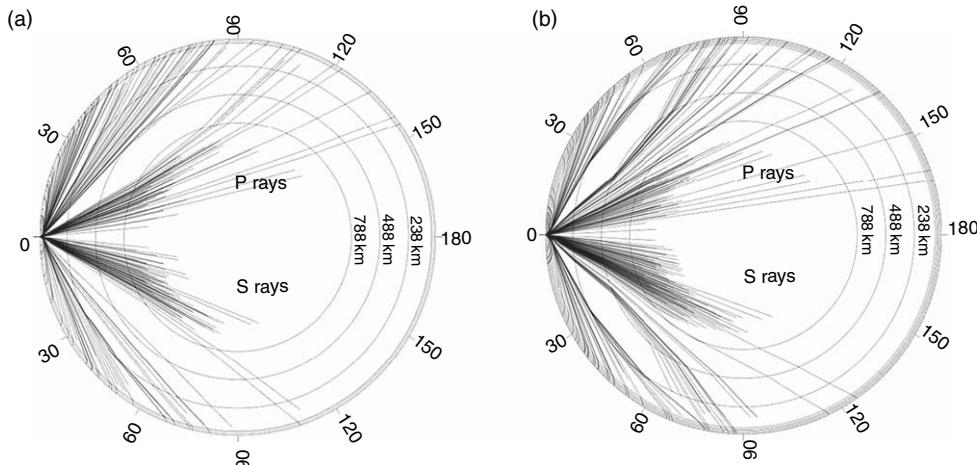


Figure 6 Summary of the source–receiver seismic ray paths for two different seismic models of the Moon determined (a) at the top with the traveltimes of Gagnepain-Beyneix *et al.* (2006) and (b) at the bottom with the traveltimes of Nakamura (1983). All pairs of rays are superimposed, with the stations all located at 0° and the source at the actual depth and epicentral distance from the station. (Note the rays associated with deep moonquakes.) The crustal thickness in (a) is 30 km, below which there is a 10 km transition zone in which there is a shear-wave gradient. Crustal thickness for the model in (b) is 58 km. Despite differences in the mantle sampling, no rays sound the lunar core for both models. Reproduced from Gagnepain-Beyneix J, Lognonné P, Chenet H, and Spohn T (2006) Seismic models of the Moon and their constraints on the mantle temperature and mineralogy. *Physics of Earth and Planetary Interiors* 159: 140–166, with permission from Elsevier.

that a 30–45 km thick crust at the Apollo 12/14 sites (KM02, L03, GB06) is preferred over previous estimates of ~60 km (N83). Third, these more recent crustal thickness estimates are consistent with estimates based on analyses of gravity and topography data (Wieczorek *et al.* (2006) and references therein). Fourth, although not discussed here, an attempt has been made to investigate geographical variations in crustal thickness using the Apollo seismic data (Chenet *et al.*, 2006). While only a few estimates are possible, they are in good agreement with relative crustal variations deduced from gravity and topography data (Chenet *et al.*, 2006).

Within the crust, estimates of seismic velocity structure are variable (Figure 5). While some discussion has been given to this in the literature, the variability among models reflects mainly the limited data set. Perhaps the most satisfactory summary of the seismic data is that it is broadly consistent with a crustal structure in which there are two major compositional layers – an upper anorthositic and lower noritic crust – consistent with inferences from gravity and topography data (Wieczorek *et al.* 2006). Seismologically, higher velocities are associated with the noritic lower crust. Some suggestions of a mid-crustal reflector have been made (e.g., the 20 km discontinuity of KM02). Furthermore, a near-surface regolith and fractured layer likely results in

scattering of seismic energy and significantly reduced seismic velocities in the upper ~1 km (GB06 and see review in Wieczorek *et al.*, 2006). Finally, a gradual increase of velocity with depth, may reflect the closing of microfractures with increasing lithostatic pressure.

Despite the variability in shallow structure, upper mantle (<300 km) velocity estimates are generally in agreement, around 7.7 km s^{-1} for P waves and 4.5 km s^{-1} for S waves (L03, N83, GB06). Mean upper mantle P-wave (S-wave) velocity estimates by KM02 are systematically higher (lower), although estimates from all the studies shown in Figure 5 agree within the uncertainties, typically ranging from 100 to 250 m s^{-1} for P velocities and up to about 100 m s^{-1} for S velocities (see Table 1 of Lognonné, 2005). Velocity estimates in the 270–1000 km region are constrained primarily by traveltime picks from deep events. A mid-mantle (300–500 km depth) low-velocity zone or negative velocity gradient was suggested by (N83); this is permitted, although not required (Figures 5–7 and following discussion). In particular, the crustal structure of N83 results in very poor ray coverage of the depth region 300–500 km (Figure 6); the tradeoff is that improved coverage of this region can be obtained (e.g., with thinner crustal thicknesses), at the expense of poor coverage at 500–800 km depth (Figure 6, and see the discussion in GB06).

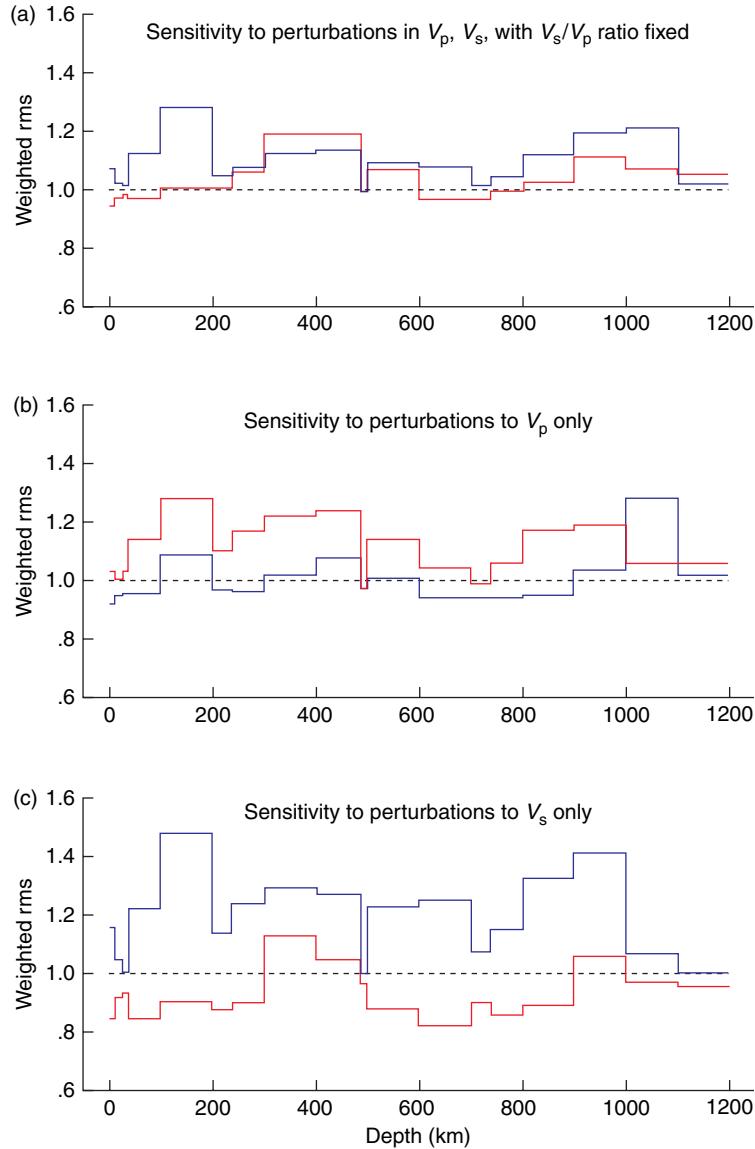


Figure 7 Weighted RMS misfit of the predicted traveltimes to observed traveltimes (Lognonné et al., 2003) for velocity models in which the velocity model of GB06 is perturbed by 10% in a specified 100 km thick interval. Where the published velocity model comprises layers of thickness less than 100 km, the published layer thickness is used. Traveltimes residuals (observed minus predicted) are weighted by the uncertainty in the traveltime pick and the total misfit is normalized such that the misfit to the unperturbed GB06 is 1 (dashed line). GB06 increased by 10% as a function of depth (red); GB06 decreased by 10% (blue). (a) the V_p/V_s ratio is held constant, as specified in GB06, the P-wave velocity is perturbed by $\pm 10\%$ in a specified depth interval and V_s is perturbed using the V_s ratio. rms misfit is computed using both P and S arrival times. (b) V_p only is perturbed by $\pm 10\%$ and the misfit to the P arrival times computed. (c) V_s only is perturbed by $\pm 10\%$ and the misfit to S arrival times computed.

Mid-mantle velocities from all three models shown in [Figure 5](#) agree within uncertainties, although differences among models affect velocity estimates below 500 km. In the depth range 500–1000 km, there is considerable difference among the models shown here – in particular estimates by KM02 are

consistently higher than those of N83, L03, and GB06. Seismic velocities below ~ 1000 km are essentially unconstrained ([Figure 6](#)). However, we note the absence of moonquakes below this depth may be consistent with possible high shear-wave attenuation, the latter having been proposed to explain the lack of

far-side deep moonquakes detection (Nakamura, 2005). Inversions of the seismic data for a lunar temperature profile (Gagnepain-Beyneix *et al.*, 2006) suggest temperatures of about 1200°C at this depth. All these observations argue for a warmer, more ductile lower mantle.

The effect on seismic velocity model resolution of raypath coverage is also illustrated by a sensitivity analysis of the model of GB06. We examine the sensitivity of the traveltimes data set to perturbations to GB06. Weighted root mean square (rms) misfits are calculated for velocity models in which the velocity is as specified in GB06, except in one depth interval. In this depth interval the velocity is either increased (red line, **Figure 7**) or decreased (blue line, **Figure 7**) by 10% relative to the published value. Depth intervals are 100 km thick intervals, except where the published model comprises a thinner layer, in which case the published layer thickness is used. The rms misfit of the traveltimes predicted by the GB06 model (**Figure 5**) to the measured traveltimes is ~ 5.9 s, comparable to the rms of the errors assigned to the data (~ 6.2 s). We instead show a weighted rms misfit (**Figure 7**) where each traveltime residual (observed minus predicted) is divided by the uncertainty in the traveltime pick (1, 3, or 10 s, see Lognonné *et al.*, 2003), and the total rms further normalized such that model GB06 fits the data to a weighted rms of 1.0.

We investigate perturbations to the P-wave and S-wave velocity models, calculating the rms misfit using only the P or S arrival times, respectively (**Figure 7(b)** and **7(c)**). In each of these cases the V_p/V_s ratio can vary compared with that specified in GB06. The traveltimes data are more sensitive to perturbations in V_s , than V_p with the corresponding rms misfits showing greater variation (**Figure 7(c)**). Models with a lower V_p/V_s ratio than GB06 (higher V_s , **Figure 7(c)**, or lower V_p , **Figure 7(b)**) compared with GB06 are not only acceptable, but in fact can result in improved misfit to the data set. (The exception is the depth interval 1000–1100 km, the source depths of some deep moonquakes.) This may seem surprising at first; however, it should be remembered that the analysis of GB06 is an inverse approach with *a priori* constraints applied, that is, the best-fit model according to some *a priori* criteria is determined; such *a priori* are not applied here. Models with higher V_p/V_s ratios (i.e., increased V_p or decreased V_s) have larger rms misfits than GB06, and with the exception of the perturbations to V_p in the depth range 500–800 km, these models result in more than a 10% rms

misfit increase, and as high as 50% for -10% perturbations to V_s in the depth range 100–200 km.

Somewhat different results are obtained if V_s and V_p are both allowed to vary, holding the V_p/V_s ratio constant. At depths shallower than 200 km, the traveltimes observations are able to discriminate among plausible models. In particular, decreases in seismic velocity at depths less than 200 km generate increases in misfit levels of more than 30%. At depths of 300–500 km, there is some sensitivity to seismic structure, models with 10% increases or decreases in velocity provide poorer fits to the data. Sensitivity at depths of 500–800 km is more limited, in particular higher velocities can be tolerated with little difference in rms misfit. Sensitivity in the region of deep moonquakes again increases; in particular, lower velocities result in a poorer fit to the data.

In summary, lunar seismic events recorded by the Apollo network have resulted in traveltimes data sets (Nakamura, 1983; Lognonné *et al.*, 2003) that enabled 1-D seismic velocity profiles to be inferred to depths of about 1200 km. Recent work indicates a revision of previous crustal thickness estimates of ~ 60 km (Nakamura, 1983) (an average value, probably most representative of the Apollo station 12 and 14 region) to less than 45 km (Khan *et al.*, 2000), about 38 km for Khan and Mosegaard (2002), and possibly as low as 30 km (Lognonné *et al.*, 2003; Gagnepain-Beyneix *et al.*, 2006). While both models are compatible with the receiver-function arrival times (Vinnik *et al.*, 2001), modeling of receiver-function amplitudes favors a thin crust with low seismic velocities. Upper-mantle seismic velocity estimates are generally in agreement. Sensitivity to seismic velocity in the mid-mantle (300 km to as deep as 700 km) is limited, and depends critically on the crustal structure; a range of models including those with small-amplitude negative seismic-velocity discontinuities/gradients at ~ 300 km depth, and those with sharp positive discontinuities at ~ 500 km depth are permitted (**Figures 6** and **7**). Seismic velocity estimates in the deep moonquake zone differ; P-wave velocities as high as 10 km s^{-1} are obtained by some of the traveltimes data inversions (Khan *et al.*, 2000; Khan and Mosegaard, 2002), and all models show higher velocities than in the upper mantle. In Section 10.03.2.4 we summarize the constraints on lunar thermal and mineralogical structure that the seismic models provide. It is likely that much of the remaining disagreement among existing 1-D velocity models results from errors in the arrival time determinations, lack of resolution, differences in the

inversion techniques used, and averaging of 3-D structure. The low resolution of the Apollo seismometers and their limited bandwidth may lead to a misinterpretation of the P and S arrival time, if the true first arrivals are too small to be detected. Differences exist among the inversion procedures used in KM02, GB06, and N83, especially in the relocalization procedure performed after each iteration. Finally, the combined effect of real 3-D structure (geographical variations in crustal thickness; spatial variations in mantle seismic velocities due to thermal and compositional heterogeneities) and poor seismic sampling of the lunar interior due to the limited source–receiver geometries means that new data provided by future networks with very broadband seismometers are required to resolve existing questions to afford major new advances.

10.03.2.3 Very Deep Interior and Joint Seismic/Gravity Inversions

The geometry of the lunar network, in particular the lack of any antipodal stations, means that few, if any ray paths propagating deep in the Moon (>1200 km depth) have been recorded (Figure 6; and see Nakamura *et al.* (1974) for an impact on the far side). An alternative approach for investigating core structure, involves investigations of a planet’s normal modes (e.g., Lognonné and Clévédé, 2002). A search for free oscillations in the Apollo data has been performed by a few authors, as low-angular-order normal modes are sensitive to core structure. After an unsuccessful attempt by Loudin and Alexander (1978), Khan and Mosegaard (2001) claimed detection of free oscillations from flat-mode LP Apollo signals generated by meteorite impacts. However, Lognonné (2005) and Gagnepain-Beyneix *et al.* (2006) have shown that the signal-to-noise ratio of these events was probably too small to result in detectable long-period signatures. Hence, seismic constraints on the lunar core are indirect. Nakamura (2005) has suggested the presence of about 30 possible deep moonquake source regions on the lunar farside; however, no events were detected within 40° of the antipode of the mean sub-Earth point, suggesting that this region is either aseismic or strongly attenuates or deflects seismic energy (Nakamura *et al.*, 1982; Nakamura, 2005).

Many geophysical studies indicate that the Moon has a core (see for a review Hood and Zuber, 2000), either based on magnetic induction signatures (Hood *et al.*, 1999), or the existence of remanent magnetism

(Hood, 1995; review in Cisowski *et al.*, 1983; Fuller *et al.*, 1987). Geochemical analyses of mare basalt samples indicate a depletion of highly siderophile elements (e.g., Righter, 2002), relative to that expected from any lunar core formation scenario (Canup and Asphaug, 2001). Impact simulations (see for a review Cameron, 2000) suggest that a low fraction of iron from the proto-earth and proto-Moon was put into orbit after the giant impact. These mass-fraction estimates are typically 1% or less, and reach 3% in only a few extreme cases. Further iron can, however, be added during late-stage accretion.

The only methods that have directly investigated the lunar core are magnetic sounding and geodesy. Magnetic sounding (Hood *et al.*, 1999) is based on the induced magnetic dipole moment produced by the motion of the Moon through the Earth’s geomagnetic tail. A core radius of 340 ± 90 km is inferred, under the assumption that electric currents in the core can be approximated by a current ‘sheet’ localized on the core surface. The second approach, measurement of the moment of inertia ratio (0.3932 ± 0.0002 ; Konopliv *et al.*, 1998), indicates a higher density toward the center of the Moon than inside the lunar mantle. Moreover, analyses of lunar rotation (Bois *et al.*, 1996; Williams *et al.*, 2001) have shown that the rotation of the Moon is influenced by a dissipation source, which has been interpreted as the signature of a liquid core.

More precise interior structure models can be obtained from joint inversion of the density, moment of inertia, Love number (k_2), and induction signature, with or without the additional constraints provided by the seismic data. Bills and Rubincam (1995) used the mean density and the inertia factor only, and estimated a core size between 400 and 600 km for densities of 8000 and 6000 kg m^{-3} , respectively. Khan *et al.* (2004) used these constraints, along with the Love number, and performed a Monte Carlo inversion assuming a five-shell model. A core with a radius of about 350 km, with a density of 7200 kg m^{-3} was inferred.

As several tradeoffs exist between the size and density of these layers, the independent constraints from seismology can be added in order to limit the space of acceptable models. Interior structure inversions based on *a priori* seismic models were first performed by Bills and Ferrari (1977), using a preliminary seismic model, and later by Kuskov and Kronrod (1998) and Kuskov *et al.* (2002), using Nakamura’s (1983) seismic model. Either a pure $\gamma\text{-Fe}$ core with density of 8100 kg m^{-3} and a radius

of 350 km, or a core with smaller densities and larger radius, including the largest troilite FeS core with a radius of 530 km and a density of 4700 kg m^{-3} were proposed. Khan *et al.* (2006) performed another study using seismic information, the inertia factor and the mean density and found a core with a density of about 5500 kg m^{-3} . Figure 8 shows the results of our inversion, where the k_2 Love number is used together with the density, inertia factor, and an *a priori* seismic model. The inverted parameters are the density of the middle and lower mantle, the shear velocity of the lower mantle, and the density and radius of the core. Both the seismic models of Nakamura (1983) and of Gagnepain-Beyneix *et al.* (2006) are used. The results show that a wide range of acceptable core models comprising 1–2% of the lunar mass fit the data. For a given mass fraction, higher S-wave velocities are necessary in the lower mantle for the Gagnepain-Beyneix *et al.* (2006) seismic model than for Nakamura (1983) model. Densities are generally less than 6000 kg m^{-3} , suggesting a core containing some light element(s). This is consistent with estimates of the temperature at the core–mantle boundary, which are compatible with a liquid core only if the latter contains light elements (Gagnepain-Beyneix *et al.*, 2006; Lognonné *et al.*, 2003; Khan *et al.*, 2006). A core with little or no light elements, corresponding to the high densities found by Khan *et al.* (2004), will likely be solid at those temperatures and can be excluded.

Finally, we note that the depths of deep moonquakes provide another important constraint on deep lunar structure. As shown above, the inversion of seismic models and gravity data allow calculation of acceptable interior models. Density and elastic moduli from these models can then be used to explore tidal stresses as a function of depth (Figure 9(c)) and/or time (Bulow *et al.*, 2006). Understanding tidal stresses as a function of time and position is critical to understanding how and why deep moonquakes occur, because the seismic data distribution and quality prohibit the inference of focal mechanisms for these events.

10.03.2.4 Mineralogical and Thermal Interpretation of Lunar Seismic Models

Considerable effort has been expended in interpreting the depth dependence of lunar seismic velocity in terms of mineralogical and thermal structure. An excellent review of this topic has been given recently by Wieczorek *et al.* (2006), and so we focus here on

only the major results and outstanding issues from these studies. Two general approaches have been used: investigations of a limited suite of compositional/mantle mineral assemblages (e.g., Hood and Jones, 1987; Mueller *et al.*, 1988; Lognonné *et al.*, 2003; Gagnepain-Beyneix *et al.*, 2006), and a more complete thermodynamic treatment (series of papers by Kuskov and co-workers, e.g., Kuskov, 1995, 1997; Kuskov and Kronrod, 1998; Kuskov *et al.*, 2002 and recent work by Johnson *et al.*, 2005; Khan *et al.*, 2006). In the latter approach, a bulk composition is specified as a function of depth and, the equilibrium mineralogical assemblage is calculated for a given selenotherm (lunar temperature profile), from which the elastic moduli and resulting seismic velocities can be predicted. Seismic velocity models predicted from mineralogy can then be compared with those inferred from traveltimes data. Alternatively, traveltimes can be calculated using the predicted seismic velocity models and compared directly with the observations (Johnson *et al.*, 2005; Khan *et al.*, 2006). Of particular interest for the Moon, have been (1) the bounds that crustal thickness (through the concentration of heat-producing elements in the crust) and mantle seismic velocities place on temperature, (2) whether mantle seismic velocities can discriminate among broad classes of interior differentiation models, and (3) whether phase changes are identifiable in the seismic velocity structure.

Studies to date (Hood and Jones, 1987; Mueller, 1988; Khan *et al.*, 2006) show that no single compositional model predicts seismic velocities that match both the upper- and lower-mantle 1-D profiles, even when temperature and pressure effects are accounted for. In general, models that assume differentiation of the upper mantle and that are more aluminous in bulk composition than the terrestrial mantle are favored. As most mineralogical studies compared their predicted seismic velocities with the Nakamura (1983) model, a focal point was the presence or absence in the predictions of a velocity discontinuity at 500 km; both phase changes (Hood and Jones, 1987; Mueller *et al.*, 1988) and compositional boundaries (Mueller *et al.*, 1988; Kuskov, 1997) at this depth were inferred. Compositional boundaries were of interest as they may reflect initial compositional zoning of the Moon, the base of and early lunar magma ocean or its resulting crystallization sequence, and/or the maximum depth of melting of the mare basalt region (see discussion in Wieczorek *et al.*, 2006). The discussion in Section

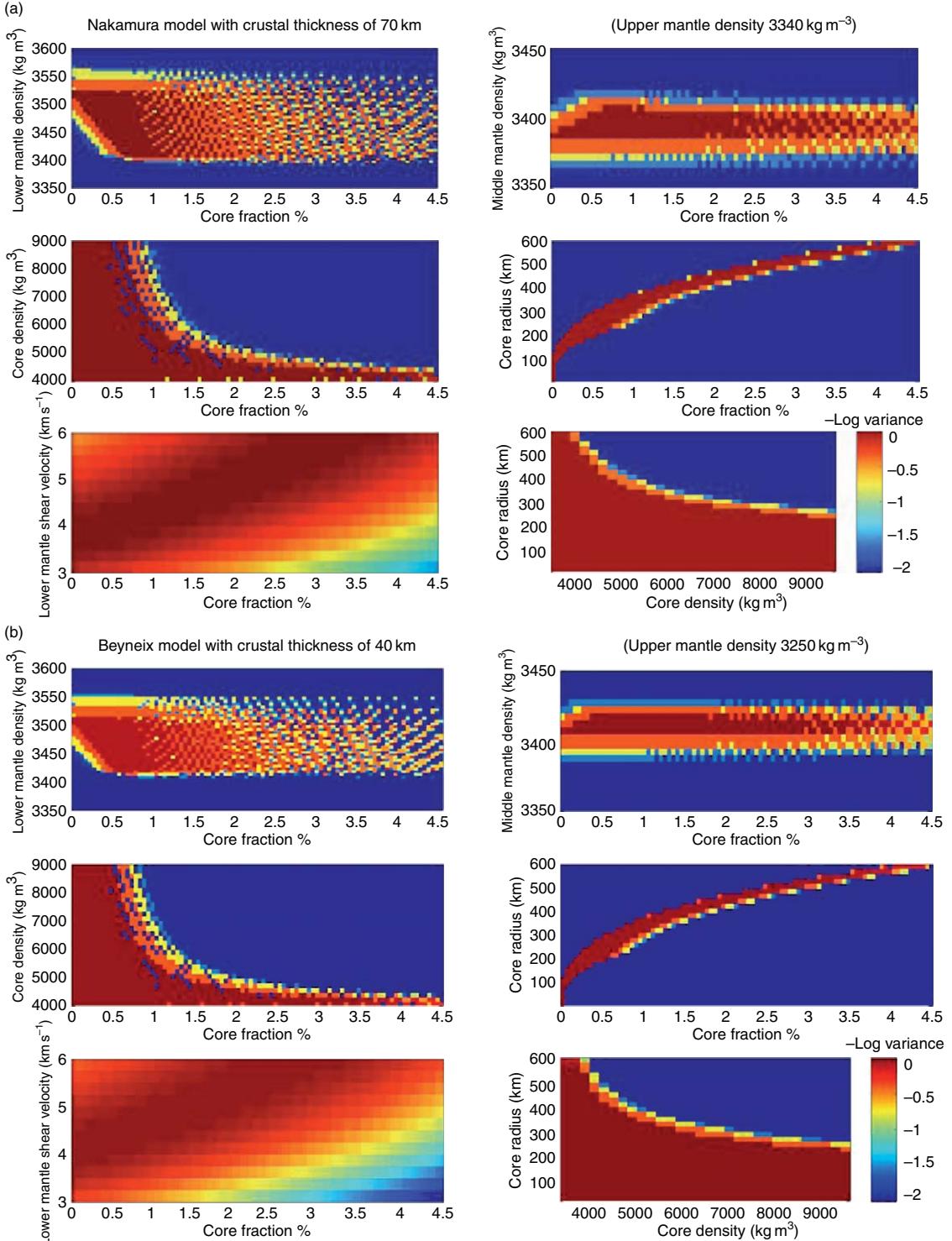


Figure 8 Exploration of the model space for acceptable models for density, inertia factor, and Love number k_2 using (b) the seismic models of Gagnepain-Beyneix *et al.* (2006) and (a) Nakamura (1983). The middle mantle is defined between 1500 and 1000 km radius, while the lower mantle is between 1000 km and the core radius. The color scale represents the decimal logarithm of $\exp(-\text{var})$, proportional to the probability, where the variance is between the computed and observed densities, moment of inertia, and k_2 . For the variance definition, values and errors, see Khan *et al.* (2004). Dark red and red are acceptable models. The model space is sampled in order to identify the range of acceptable solutions. In the middle/lower mantle S-wave velocities values are equal to 4.65 km s^{-1} and 4.5 km s^{-1} in the models of Nakamura (1983) and Gagnepain-Beyneix *et al.* (2006), respectively.

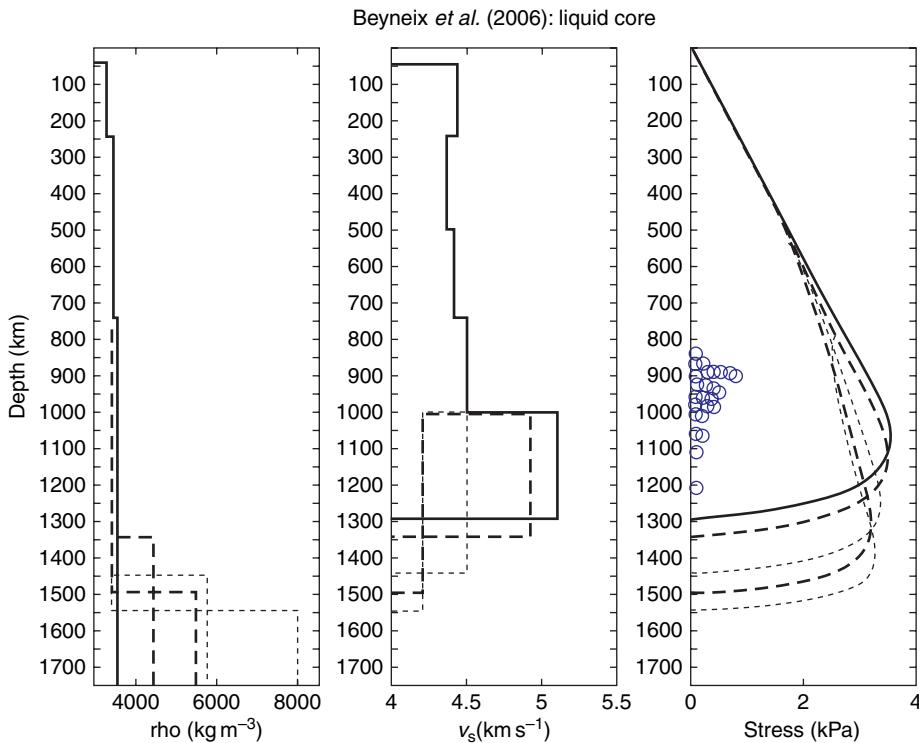


Figure 9 Several seismic and density models of the Moon matching the Apollo seismic traveltimes, the mean density, moment inertia and k_2 within the data error bars. All models have S-wave velocity values of Gagnepain-Beyneix *et al.* (2006) for the mantle and crust and only the shear velocity in the very deep mantle is modified. Shear velocity in the core is zero, as only models with liquid core are shown. The various lines are each associated with a given core size and minimize the variance. Figures from left to right represent the density, the shear-wave velocity, and the tidal stresses. Models with the largest core (400 km or more) correspond to an ilmenite core (Wieczorek and Zuber, 2002) with core densities lower than 5000 kg m^{-3} . These have correspondingly high shear velocity in the lower mantle. Models with a core radius of 350 km correspond to an FeS core, with densities in the range of $5000\text{--}6000 \text{ kg m}^{-3}$. Smaller cores ($\sim 200 \text{ km}$) of pure Fe are also compatible with the data if associated with a low-velocity zone in the lower mantle, in order to match the low k_2 value. On the far right, the maximum horizontal tidal stress, with respect to depth, is shown, defined as $(T_{\theta\theta} + T_{\phi\phi})/2$, where T is the tidal stress tensor, at the latitude and longitude of the deep Moonquake A1 (as found by Gagnepain-Beyneix *et al.* (2006), i.e., -15.27° S , -34.04° E). See Minshull and Goult (1987) for more details on stress computations. Note that only models with a core radius of 350 km or more produce maximum tidal stresses in the vicinity of the deep moonquakes.

10.03.2.2 indicates limited sensitivity of the arrival times to velocity structure in the region 300–500 km, and as deep as 800 km, and cautions against over-interpretation of mantle structure in this region. Overall, however, the increase in seismic velocity from the uppermost mantle (0–300 km) to the region of deep moonquakes (700–1200 km) is consistent with a change in bulk composition from a dominantly orthopyroxene upper mantle (with lesser amounts of olivine, clinopyroxene, plagioclase, and garnet) to a dominantly olivine lower mantle (with smaller amounts of garnet and clinopyroxene) (Wieczorek *et al.*, 2006). Associated with such a compositional change is an increase in magnesium number with depth. The recent studies by Lognonné *et al.* (2003), Gagnepain-Beyneix *et al.* (2006), and Khan *et al.*

(2006) demonstrate the broad range of acceptable models, given current seismic constrains.

Crustal thickness plays an important role in determining the thermal profile of the lunar interior due to the preferential sequestration of heat-producing elements into the crust. Again a large range of thermal profiles and resulting mineralogical assemblages are compatible with the seismic data (Khan *et al.*, 2006); recent studies indicate that ‘cold’ thermal profiles are preferred (Lognonné *et al.*, 2003; Gagnepain-Beyneix *et al.*, 2006), with temperatures of 1073 K (equivalent to the base of the terrestrial elastic lithosphere) and 1473 K (base of terrestrial thermal lithosphere) at depths of 340 km and 740 km, respectively. Such depths are comparable to the depth found in thermal evolution models of the

Moon (e.g., Spohn *et al.*, 2001a). Bulk lunar abundances of the heat-producing elements uranium and thorium in these models are similar to terrestrial values, with an enriched crust and a depleted mantle. For example, mantle abundances of about 8.2 ppb of U and 30 ppb of Th are suggested by Gagnepain-Beyneix *et al.* (2006) in order to fit the temperature dependence of seismic velocity. These values are close to those proposed by Waenke *et al.* (1977) and Taylor (1982).

Taken together, mineralogical and thermal studies of the lunar interior that use seismic data show that a broad range of interior models are compatible with the data. Limitations of existing models include the details of the thermodynamic parametrization, and the absence of titanium (and sometimes sodium) from calculations. Furthermore, the likelihood of 3-D structure in the Moon is great, and the broad range of acceptable mineralogical and thermal models may simply reflect 3-D structure in the seismic data that has been (necessarily, given the data set) mapped into 1-D structure.

10.03.3 Seismic Activity of the Moon and Terrestrial Planets

10.03.3.1 Internal Seismic Activity

So far, no conclusive indication of present-day plate tectonic activity has been observed on a planet other than Earth. Moreover, only the seismic activity of the Moon is constrained by seismic data and from a practical perspective only the activity of Mars and Venus can be estimated. During the 7 years of the Apollo seismic network operation, about 12 500 seismic signals were detected on the LP instruments and cataloged, and many more events seen on the SP instruments remain uncataloged. Figure 10 shows the statistics of detection on the horizontal component of Apollo 14 station per year, based on Nakamura's catalog. We focus first on the internal activity of the Moon, then address the internal activity of other planets, followed by a discussion of impact seismology.

Assuming that the largest deep moonquakes have a seismic moment of 5×10^{13} N m and associated stress drops of about 10 kPa (Goins *et al.*, 1981a), quakes with seismic moment 30 times smaller were detected. These smallest reported moonquakes correspond to terrestrial events with body-wave magnitudes as low as 1.6 and are detectable due to the low seismic noise level on the Moon (Figure 1).

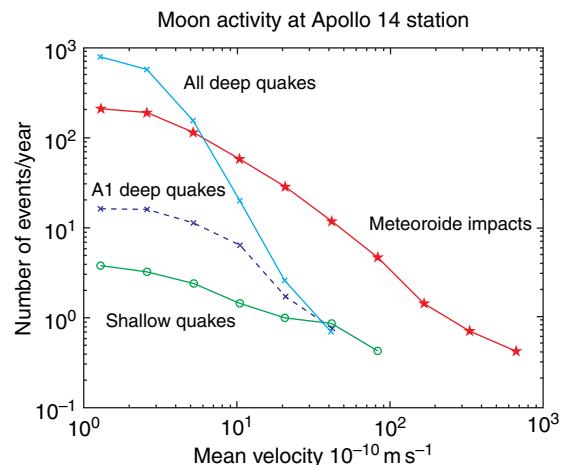


Figure 10 Number of events/year detected on the Apollo 14 station vs amplitude. Amplitudes are from the Nakamura *et al.* (2004) catalog and converted approximately to mean velocity, assuming a recording by the long-period horizontal component on the peaked mode and peaked frequency. Note the very high number of events detected, despite their very low amplitudes (see Figure 1 for seismometer sensitivity).

Note however that the definition of the body-waves magnitude or surface-wave magnitudes are generally defined for the Earth. We will therefore use later, unless specified differently, moment magnitudes, directly related to the seismic moment by

$$M_w = 2/3(\log_{10}(M_0/Nm) - 9.1)$$

where M_w is the moment magnitude and M_0 the seismic moment in N m (Hanks and Kanamori, 1979).

Deep moonquakes originate from regions that appear to undergo repeated failure, giving rise to sets of moonquakes with similar waveforms and periodic occurrence times (Lammlein *et al.*, 1974). The number of known source regions for deep moonquakes is currently estimated as ~ 250 (Nakamura, 2003, 2005), and source depths are 700–1200 km. Figure 11(a) shows the variation in the number of moonquakes recorded per week, for the duration of the Apollo seismic experiment. The installation dates of the stations are provided in Table 1. The activity at all known moonquake nests (reported in Nakamura *et al.*, 2004) is shown in red, and the activity recorded at nine clusters that dominate the catalog is shown in blue. Figure 11(a) illustrates that activity at these nine well-studied clusters closely resembles the behavior of the larger deep moonquakes population. Individual peaks in the time series occur at approximately 2-week and 4-week intervals. More obvious over the duration of the experiment is the modulation of

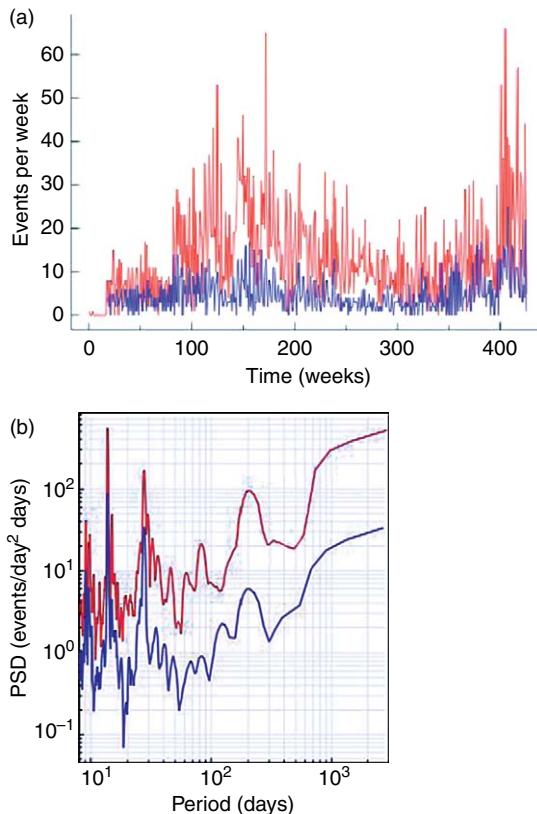


Figure 11 Deep moonquake activity. (a) Number of moonquakes per week vs time in weeks, starting from the beginning of the seismic experiment is shown. Moonquake activity for a recent update of the deep moonquake catalog (Nakamura et al., 2004), is shown in red. For comparison, moonquake activity at nine of the most active deep clusters is shown in blue. This includes new events recently found by Bulow et al. (2005, 2006) and not included in the lunar catalog (Nakamura et al., 2004). (b) Power spectral density of number of moonquakes per day vs period in days. Peaks at \sim 14, 28, and 206 days can be seen, and power at the longest periods likely results from a 6-year modulation of moonquake occurrence times. Red and blue curves are as in (a).

moonquake activity at the 206-day period, and possibly at a period close to the duration of the experiment (6 years). **Figure 11(b)** shows the power spectrum of moonquake activity. The number of moonquakes per 24-h day was calculated, and the power spectrum computed using the minimum-bias multitaper method of Riedel and Sidorenko (1995). A multitaper method was chosen because tidal effects with periods of 206 days and 6 years are suggested by the time series in **Figure 11(a)**. The 6-year period in particular, requires the full length of the time series (8 years) to be analyzed. Various choices of number of tapers were explored; **Figure 11(b)** shows the power spectrum

computed using six tapers. This number allows frequency resolution at the expense of noisy spectral estimates, as is evident at high frequencies. Spectral peaks are seen at 13–14 days and 27–28 days, with some structure seen in the monthly peak. A broad peak in the power spectrum is seen close to the previously noted 206-day period, and the increase in power at the longest periods likely reflects the 6-year modulation of moonquake occurrence times (Lammlein et al., 1974; Lammlein, 1977) (Energy from this period is spread over the five lowest-frequency spectral estimates which occur at periods of 2916, 1458, 972, 729, and 583 days). Increased smoothing using more tapers or the fully adaptive multitaper approach results in spectra with broader peaks at the monthly and fortnightly periods, and the loss of any resolution at around 200 days. Thus, this spectral approach, which is typically used in investigating tidal periodicities in terrestrial quakes (e.g., Tolstoy et al., 2002), provides only crude insight into the lunar data.

In addition to the deep quakes, 28 other events were detected at much shallower depths. These events may be analogous to terrestrial earthquakes. They show no obvious correlation with tidal activity, nor with surface features on the Moon. No events in the southeast quadrant of the nearside of the Moon were detected, and focal depths are below the crust, but above 200 km depth. The largest shallow events have a seismic moment of about 3×10^{14} N m for Goins et al. (1981a) or 1.6×10^{15} N m for Oberst (1987). On Earth, events with such moments will have body-wave magnitudes ranging from 4.8 to 5.5. Their depths have been estimated to be less than 200 km, and they are interpreted as resulting from the release of tectonic stresses (Nakamura et al., 1982). Flat displacement spectra were recorded even on the SP seismometers, indicating a corner frequency higher than 10 Hz in most the cases (Goins et al., 1981a). This implies very high stress drops – less than 40 MPa for Goins et al. (1981a) but up to 210 MPa for Oberst (1987) – perhaps due to the cold, volatile-poor (and hence rigid) lunar lithosphere. Interestingly, these events are similar to Earth’s intraplate earthquakes, which also exhibit much higher stress drops (for a given seismic moment) than plate tectonic earthquakes (Scholz, 1990). (Intraplate earthquakes have been used a reference for all estimates of the seismic activity of telluric planets.)

We now examine how quakes can be released by the thermoelastic cooling of the lithosphere. We apply the theory to the Moon, in order to possibly understand the shallow lunar events, and to Mars, in order to predict a

possible background seismic activity. The cumulative seismic moment, over a time Δt , is given by

$$M_{\text{cum}} = \eta \dot{\varepsilon} V \mu \Delta t \quad [1]$$

where V is the seismogenic volume, η is the seismic efficiency, and μ is the mean shear modulus of the seismogenic layer (Phillips, 1991; Bratt *et al.*, 1985).

Following Phillips (1991), the strain rate $\dot{\varepsilon}$ is given as

$$\dot{\varepsilon} = \frac{1}{H(a - \frac{H}{2})^2} \int_{a-H}^a \alpha \dot{T} r^2 dr \quad [2]$$

where H is the thickness of the seismogenic layer, α is the thermal expansion coefficient, \dot{T} is the cooling rate, and a is the planetary radius. If we assume that the thermal gradient in the seismogenic layer is linear and given by

$$T = T_0 + ((a - r)/H)\Delta T \quad [3]$$

and that H thickens with time at a rate \dot{H} (both \dot{T} and \dot{H} are counted positively), we then have

$$M_{\text{cum}} = 2\pi\alpha\eta\mu a^2 H \Delta T \left(\frac{\Delta \dot{T}}{\Delta T} + \frac{\dot{H}}{H} \right) \beta \Delta t \quad [4]$$

where ΔT is the contrast of temperature between the bottom of the seismogenic layer and the surface and where

$$\beta = \left(1 - \frac{H}{a} + \frac{1}{3} \frac{H^2}{a^2} \right) \frac{1 - (4/3)(H/a) + (1/2)(H^2/a^2)}{(1 - (H/a)) + ((1/4)(H^2/a^2))} \quad [5]$$

is a geometrical factor correcting the thin-shell approximation. It is more than 0.95 for Mars, the Earth or Venus and 0.72 for the Moon.

For the Moon, we have assumed $H=400$ km, an estimate of the depth of the 1073 K isotherm (Spohn *et al.*, 2001a). The isotherm corresponding to the base of the lunar seismogenic layer is unknown. Abercrombie and Ekström (2001) take a 873 K for the limit of brittle failure in oceanic crust, which is reached at about 250–300 km on the Moon, using the thermal models of Spohn *et al.* (2001a) and Gagnepain-Beyneix *et al.* (2006), respectively. None of the shallow moonquakes occurred at depths greater than 200 km, however. On the other hand, the lunar crust might be rich in anorthite and in that case, the pressure and temperature at that depth might be those of the transition of feldspath from brittle to plastic, found between 20 and 30 km in the Earth's continental crust (Scholz, 1990). Deep moonquakes occur at much larger depth, up to 1100–1200 km, where the temperature seems high

(1600–1700 K) for all published models. This remains a paradox, either related to our models of the lunar interior, lunar mineralogy, or to our current understanding of fault rupture under these conditions. However, as noted above, the deep moonquakes release limited energy compared with the shallow quakes.

For Mars, Phillips (1991) has assumed a constant cooling rate of \dot{T} equal to 1.1×10^{-7} K yr⁻¹, and using a thermal evolution model Knapmeyer *et al.* (2006) have set the \dot{T} value to 0.5×10^{-7} K yr⁻¹. Taking the 873 K lower bound on the isotherm at the base of the seismogenic layer, and the thermal evolution models of Spohn *et al.* (2001b), we estimate \dot{H} in the range 3.15×10^{-8} to 4×10^{-8} km yr⁻¹ (For comparison, this is 10–20 times smaller than the rate of thickening of 100 Ma oceanic lithosphere.) Using $\alpha = 2 \times 10^{-5}$ K⁻¹, $\mu = 66$ GPa (obtained for $V_s = 4.5$ km s⁻¹ and $\rho = 3300$ kg m⁻³) and 3.5×10^{-8} km yr⁻¹ for \dot{H} , $M_{\text{cum/year}} = \alpha\mu\Delta t 2\pi a^2 \dot{H} \Delta T \approx 5.5 \times 10^{17}$ Nm yr⁻¹, a value to be compared with the annual release of the shallow quakes (7.3×10^{14} Nm yr⁻¹). Such a rate is comparable to the one obtained if a constant $\dot{T} = 0.5 \times 10^{-7}$ K yr⁻¹ is taken for a 200 km seismogenic layer. The ratio between observation and model is about 1.3×10^{-3} , 20 times smaller than the lowest seismic efficiency parameters η reported for Earth (e.g., Ward, 1998). The shallow seismic energy release of the Moon, is then, an outstanding puzzle: calculations of the type above, modeled on the approach of Phillips (1991) are either inappropriate or require extreme choice of parameters relative to our current understanding. Lunar seismic efficiency may be lower than on Earth, possibly due to a cold, volatileless seismogenic layer able to support large initial stress without rupture. On the other hand, if these calculations are correct, we cannot exclude much larger shallow moonquakes, with a recurrent time however larger than the 7 years of operation of the Lunar Network.

The activity of Mars is unknown and can be estimated only by a comparative approach with respect to the Moon, as shown on Figure 12. Only one event has been tentatively proposed as a quake, during the 19 months of nearly continuous operation of the Viking Lander 2 seismometer (Anderson *et al.*, 1977a, 1977b). During most of the experiment, the seismometer signal was correlated with the wind-related vibrations and lift of the lander. (See Lognonné and Mosser (1993) for a detailed explanation, and Nakamura and Anderson (1979) for an application of the seismometer as wind sensor.) As the proposed event was detected at

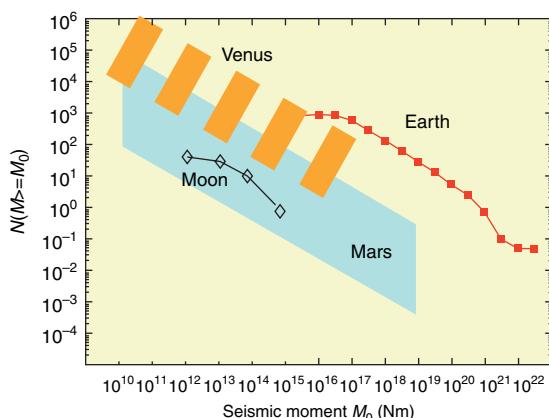


Figure 12 Seismic activity of the telluric bodies. The figure shows the number of events per year larger than a given seismic moment. For the Moon, the estimate is obtained from Oberst (1987) and is only for the shallow moonquakes. For Mars and Venus, the activity is estimated from various published models, based on the thermoelastic cooling of the lithosphere. The terrestrial activity is the mean activity in the period 1984–2004. Adapted from Knapmeyer M, Oberst J, Hauber E, Wählisch M, Deuchler C, and Wagner R (2006) Implications of the Martian surface fault distribution and lithospheric cooling for seismicity: A working model. *Journal of Geophysical Research* 111: E11006 (doi:10.1029/2006JE002708).

a time when no meteorological wind data were recorded, this single detection is far from conclusive! Moreover, Goins and Lazarewicz (1979) have shown that the 4 Hz frequency of the seismometer, due to the Martian *a priori* mantle attenuation, was not optimized for the detection of remote events and that a seismic activity of Mars comparable or lower than the terrestrial intraplate activity was still compatible with this lack of detection.

An activity lower than the Earth's one is indeed found by all the available theoretical predictions. Golombek *et al.* (1992) use surface fault observations, and Phillips (1991) use a lithospheric thermoelastic cooling model to propose an activity 100 times greater than the shallow moonquake activity detected by the Apollo seismometers. The greater activity results from the effect of surface area (a factor of about 4), a seismogenic efficiency closer to Earth values (a factor of at least 20), and a possibly greater cooling rate (a factor of about 2). However, amplitudes of the largest Mars seismic signals are still expected to be about four orders of magnitude lower than those of the largest earthquakes at long periods, that is, at frequencies below the source cut-off. Figure 13 shows that Martian seismicity might result in about 50 quakes with a seismic moment

greater than 10^{15} Nm, with an increase/decrease of the quake frequency by a factor of five for a decrease/increase of the seismic moment by a factor of 10. This represents a cumulative seismic energy release of 10^{18} – 10^{19} Nm yr $^{-1}$ (purple curve, Figure 13). These early estimates have been updated by Knapmeyer *et al.* (2006), with seismic activity ranging from a high seismic moment budget distributed over many events, to a low seismic moment budget distributed over a few events (Figure 13).

Other sources of seismicity, not modeled here, are possible. Volcanic activity, in the form of volcanic tremors or magma chamber retreat and associated faulting (e.g., as observed on Alba Patera by Cailleau *et al.*, 2003) could lead to quakes, but although surface evidence for activity in the last 10 My has been proposed (Neukum *et al.*, 2004), no evidence for present activity exists. The cooling of these most recent volcanoes along with stresses associated with known gravity anomalies (e.g., at Argyre or Isidis (Zuber *et al.*, 2000)) can supplement thermoelastic stresses. Landslides can also be associated with locally increased seismic activity. The distribution of potential tectonic quakes was studied by Knapmeyer *et al.* (2006) using the MOLA altimetry data. A total of about 7000 faults with a cumulative length of 600 000 km were found, half of them thrust faults and the other half-normal, with no obvious correlation between fault density and ages. Figure 14 shows the results of a Monte Carlo simulation of seismic activity on these faults, and shows that activity in the Tharsis region is very likely (Knapmeyer *et al.*, 2006).

We mention briefly seismicity estimates for two other planets that might be targeted for seismic missions in the coming decades: Venus and the Jovian satellite Europa. Estimation of the seismic potential activity of Venus, performed as part of the NASA Venus Internal Structure Mission study (Stofan *et al.*, 1993), suggested a seismically active planet. Assuming a seismogenic layer of 30 km, more than 100 quakes of M_w greater than five could be released by an intraplate activity with a strain rate of 10^{-19} s $^{-1}$ (Grimm and Hess, 1997). Quakes with M_w greater than six might be 5 times less frequent. A suggested rise in surface temperature on Venus, over the period for which we have a geological record, generates compressive thermoelastic stresses in the crust (Solomon *et al.*, 1999; Dragoni and Piombo, 2003). Such stresses might either generate reverse faults, or act on pre-existing reverse faults, and, by analogy with Earth, might lead to quakes of maximum moment magnitude 6.5. A complicating issue on

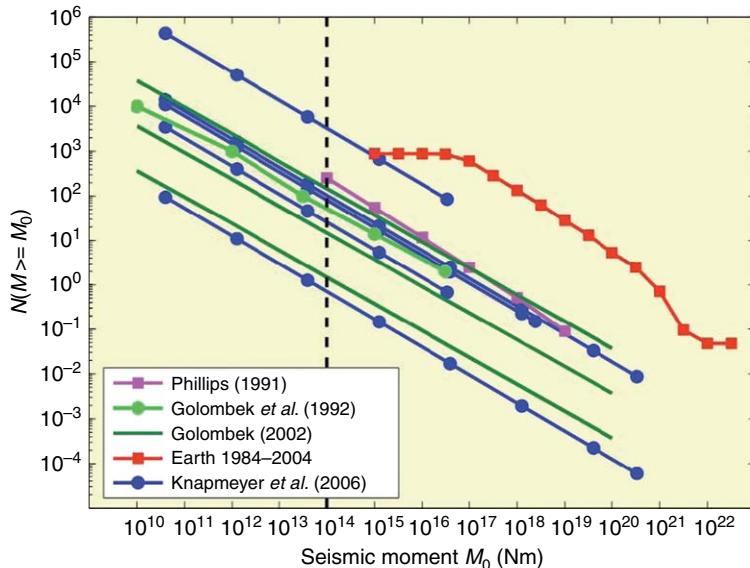


Figure 13 Comparison of models for Martian seismic activity per year with the terrestrial mean annual activity (Knapmeyer, personnal communication). Most models predict a total of 10–100 quakes per year with moments larger than 10^{15} Nm. The dashed line is for quakes of moment 10^{14} , for which S and P body waves have peak-to-peak amplitudes of about 10^{-9} m s $^{-2}$ and 2×10^{-8} m s $^{-2}$ respectively, in the 0.1–1 Hz frequency band and at 100° epicentral distance (Mocquet, 1998). This moment may therefore be at the limit of detection for estimated Martian noise levels (Lognonné et al., 2000). The multiple lines shown for Golombek (2002) and Knapmeyer (2006) are for different hypotheses regarding the seismicity.

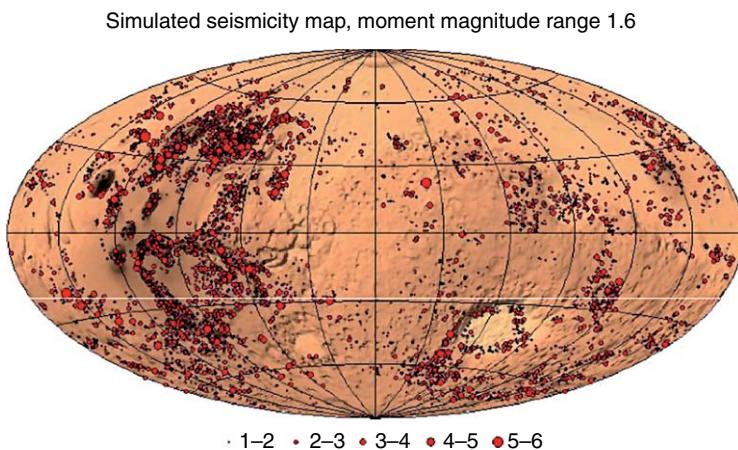


Figure 14 Seismicity map of Mars from Knapmeyer (2006), determined from a Monte Carlo simulation of seismic activity releasing a cumulated moment of 10^{18} – 10^{19} Nm per year (from Golombek et al., 1992), and with a probability distribution of epicenter locations provided by the density of surface faulting. The map shows the distribution of Marsquakes that might be expected during one Earth year, with an annual seismic budget between the upper and lower bounds (Knapmeyer, personal communication).

Venus is that of rheology; the absence of volatiles means that crustal rocks are much stronger than their terrestrial counterparts (Mackwell et al., 1998). Finally, as we note in our concluding remarks, there are severe technical challenges to surface-based seismological experiments on Venus.

The seismic activity of the Jovian satellites is of course also unknown, although Io could be the most active of all known satellites in the solar system (Kuskov et al., 2000) and a perfect target for a seismological mission. The possible activity of Europa has been reviewed by Panning et al. (2006), following

studies of Kovach and Chyba (2001) and Lee *et al.* (2003) on the possibility of seismic experiments on this satellite of Jupiter. Putative seismic sources are the opening of tensile cracks due to the Jovian tides; these might lead to quakes ranging from 2 to 4 in moment magnitude. Nimmo and Schenk (2005) have also identified regions of normal faults from an analysis of Galileo data, indicating potential quakes up to possibly $M_w = 5.3$ magnitude. The frequency of such proposed events remains unknown.

10.03.3.2 External Seismic Activity: Artificial and Natural Impacts

Impacts constitute about one-fifth of the detected events (1742 of the 9315 identified and classified events) on the Apollo seismic network, and the strongest ones have amplitudes comparable to the largest moonquakes. As discussed in Section 10.03.2, these surface seismic sources permit more detailed studies of crustal seismic structure, including lateral variations

in thickness (Chenet *et al.*, 2006). A meteorite impact can be modeled as a point force seismic source, with an amplitude equal to the momentum of the impactor (i.e., mass times impact velocity, in kg m s^{-1} or Ns). During the Apollo seismic experiment, controlled impacts of the Saturn V upper stages and the Lunar Modules were performed, and it may be possible to use these to calibrate the energy flux from natural impacts. Signals from the impact of the Apollo 17 upper stage are shown in Figure 15. The momentum of this impact was about $37\,000 \text{ kg km s}^{-1}$ ($14\,487 \text{ kg}$ at 2.55 km s^{-1}), and knowledge of its position and occurrence time, provides unique seismic data with well-constrained P- and S-wave arrival times.

Even if the locations of natural impactors are not known, they are still useful seismic sources. In the case of the Moon and Earth, the frequency and size of impactors can be obtained using statistical properties of the Earth-crossing asteroids (ECAs, e.g., Poveda *et al.*, 1999) and of the collision rates (e.g., Shoemaker *et al.*, 1990). Figure 16 shows the frequency

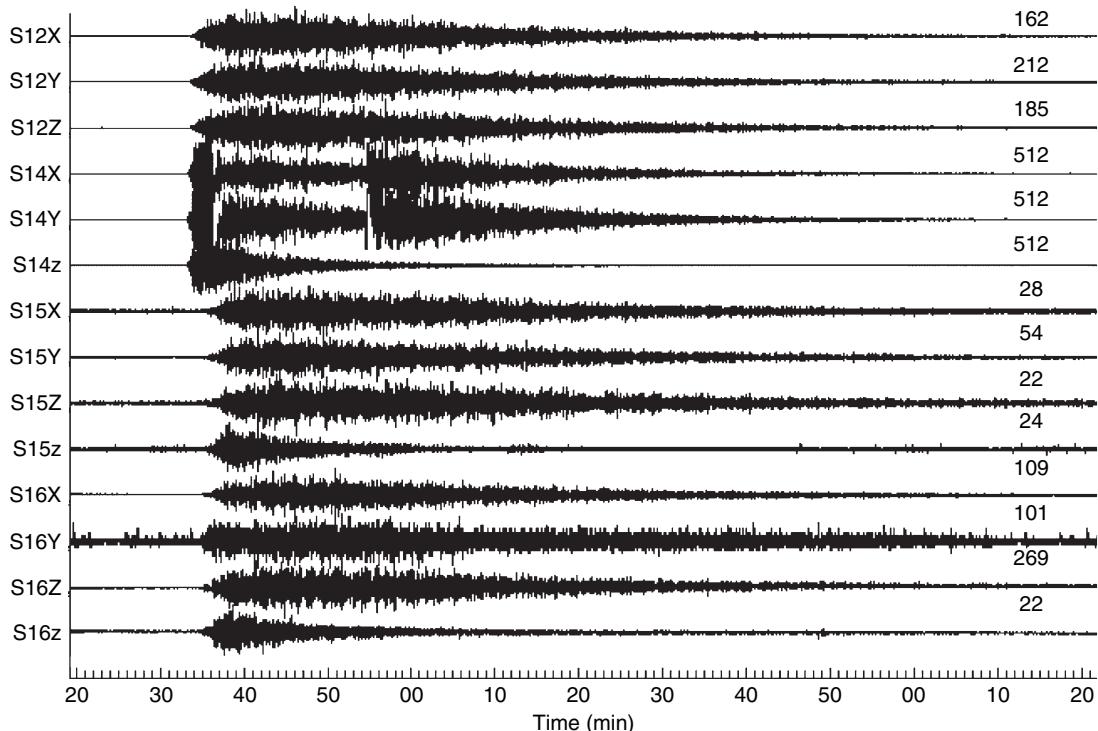


Figure 15 Seismic records, from the Apollo seismic network, of the impact of the Apollo 17 Saturn V upper stage (Saturn IVB) on the Moon on 10 Dec 1972 at 4.21°S , 12.31°W (at distances of 338, 157, 1032, and 850 km from the Apollo 12, 14, 15, and 16 stations, respectively). X, Y, Z are for the long-period seismometers, z is for the short-period seismometer. Bit sensitivity is $4 \times 10^{-10} \text{ m s}^{-2}$ for the LP (at 0.45 Hz) and $1.3 \times 10^{-7} \text{ m s}^{-2}$ for the SP (at 8 Hz). Amplitudes at the Apollo 14 station, 157 km from impact, reach about 10^{-5} m s^{-2} with an amplitude mainly related to S waves trapped in the regolith. The first P arrival is typically 10 times smaller. Note the 10 db gain change at the middle of the fourth and fifth traces. Reprinted from Lognonné P (2005) Planetary seismology. Annual Review of Earth and Planetary Sciences 33: 19.1–19.34. (doi:10.1146/annurev.earth.33.092203.122605).

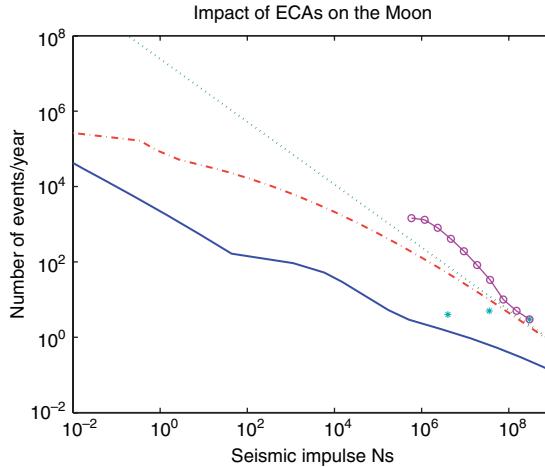


Figure 16 Impact statistics for Earth-crossing asteroids on the Moon. The distribution of impactors is from Poveda *et al.* (1999) and the collision probability is assumed to be 4.2×10^{-9} per year (Shoemaker *et al.*, 1990). Lunar case (green dotted line); impactor detection curves at the Apollo 14 station (purple circles and line), with the largest event anchored to a seismic moment of 3×10^8 N s; Mars (red dot-dashed) and Earth (blue solid) curves, with atmospheric effects included. The latter leads to a reduction of both the mass and velocity of the impactor. Pale blue stars are the seismic impulses, in increasing impulse, of the LEM, the Saturn V, and the largest natural impacts, respectively. The natural impact corresponds to typical impacts expected on Near-Earth objects (NEOs) missions. Note that event frequency scales with the squared ratio of the NEO radius (crossing diameter) and the Moon's radius. At 1 AU, the number of events, compared to the Moon, are reduced by about 120 000 for a 5 km radius NEO. Entry velocity is taken as 20 km s^{-1} , for impactor mass of 2400 kg m^{-3} .

distribution of impactors as a function of seismic impulse, for the lunar cross-section. Even when uncorrected for geometrical spreading of seismic waves, the frequency of the detected impacts with respect to the amplitude of the seismic signal (extracted from Figure 1), matches the distribution of impactors fairly well for the largest impacts (note that only the distribution in impactor diameter is taken into account). The amplitude of the largest impacts was estimated by Gagnepain-Beyneix *et al.* (2006) and typically releases a seismic momentum of about 3×10^8 N s. This is a factor of 10 larger than the Apollo Saturn V upper-stage impact, and a factor 100 larger than the Lunar Module impact. We have used this value to anchor the observed impact curve on Figure 16.

On planets with an atmosphere, both the velocity and the mass of the impactors are reduced during atmospheric entry. This effect can be assessed by

integrating the impactor equation in the atmosphere (Chyba, 1993; Poveda *et al.*, 1999):

$$\begin{aligned} m \frac{dv}{dt} &= \frac{1}{2} C_D \rho v^2 A - mg \\ \frac{dm}{dt} &= \frac{1}{2} \frac{C_H}{Q} \rho v^3 A \end{aligned} \quad [6]$$

where m is the mass of the meteoroid, v its velocity, ρ its density, g the gravity, C_D the drag coefficient (here taken as 1.7, the cylinder value), C_H the heat transfer coefficient (taken as 0.1), Q the heat of ablation of the meteoroid (taken as 10^7 J kg^{-1}), A its cross section, and t is time. Figure 16 gives the effect of the Mars and Earth atmospheres on the ECAs or Moon-impacting meteoroids, for an asteroid density of 2400 kg m^{-3} , an entry velocity of 20 km s^{-1} , and standard atmospheric models (US standard atmosphere 1976) of the planets. We see that in contrast to Earth, the atmospheric shielding effect on Mars is small. Shoemaker *et al.* (1990) and Davis (1993) have proposed that the entry flux of meteorites entering the atmosphere of Mars is 2.6 times that on Earth due to the proximity of the asteroid belt. Based on these data, the number of impacts per year detected by a seismic station on Mars is comparable to that on the Moon (Davis, 1993), assuming an instrument detection level similar to that of the Apollo seismometers.

Lastly, we consider small bodies. Currently the experiment SESAME/CASSE is onboard the Rosetta mission to comet 67P/Churyumov-Gerasimenko (Kochan *et al.*, 2000), for a landing expected in 2014. This experiment aims to investigate the outermost surface of the comet by means of active and passive acoustic-wave monitoring in a frequency range from a few hundred to several kilohertz, and may reopen the study of small-bodies seismology after the failure of acoustic sensors onboard the Phobos 1 and 2 landers (Surkov, 1990). Both internal (related to degassing events) and external sources might generate seismic signals on comets, in contrary to asteroids, for which only impacts are expected. Future missions may be targeted toward asteroids, either to study the interior or as a deflection strategy in asteroid mitigation (Ball *et al.*, 2003). Both active seismology (using explosives or impactors) or passive seismology (with natural impacts) have been already proposed (e.g., Walker and Huebner, 2004). Due to their limited size, these bodies have low gravity. As a consequence, minor impacts on asteroids produce ground accelerations higher than the local gravity, and seismic

reverberation may have a major effect on the morphology of the asteroid regolith, including in the shape and density of small craters (Richardson *et al.*, 2004). The amplitudes of the body P and S seismic waves generated by impacts can be estimated with simple models. Scattering processes will affect body-wave amplitudes, and the irregular asteroid surface will strongly affect the amplitudes of surface waves.

For large impacts, such as the artificial impacts made in the Deep Impact mission (3.8×10^6 N s by the impact of 370 kg at 10.2 km s^{-1}) (A'Hearn *et al.*, 2005) or those expected in the Don Quichotte mission (4×10^6 N s by 400 kg at 10 km s^{-1}) (e.g., Ball *et al.*, 2003), the displacement amplitude of body waves can be roughly expressed by

$$a_p = \frac{F(t - (r/c_p))}{4\pi\rho c_p^2 r} \quad \text{and} \quad a_s = \frac{F(t - (r/c_s))}{4\pi\rho c_s^2 r} \quad [7]$$

where c_p and c_s are P and S velocities, ρ is bulk density, and r is radial distance (Aki and Richards, 1980) and F are the point force source functions (in N). By scaling the impact probability of Figure 16 by the ratio of the surface body areas, we see that these impacts, on a 5-km radius asteroid, will occur naturally at a rate of about one event every 1000–1500 years for an ECA flux, and will generate accelerations in the range of 0.01 – 1 m s^{-2} depending on frequency (Figure 17). As the frequency of impacts is typically a function of the diameter with a power of $-5/2$ (e.g., Poveda *et al.*, 1999), and therefore of the mass with a power of $-5/6$, accelerations 10^6 times smaller (in the range of

1 – 100 ng) will occur at a rate of 10^5 events a year at 1 AU, and possibly higher at larger distances from the Sun. This translates into rates of hundreds of events per year that could possibly be recorded by space-qualified short-period seismometers with nanogram sensitivities. Such natural impacts might provide excellent seismic sources for tomography studies of asteroids.

10.03.4 Atmospheric Seismology

10.03.4.1 Theoretical Background

Over the last decade, many observations in terrestrial seismology have shown that the coupling between the atmosphere and the solid part of the Earth leads to an exchange of seismic energy between the two subsystems. A first example of atmospheric–solid body coupling is the detection in the ionosphere of seismic signals. As shown by Lognonné *et al.* (1998a), a small fraction of the seismic energy escapes indeed into the atmosphere at frequencies higher than 3.7 – 4.4 mHz . Near the seismic source, the acoustic-generated waves propagate as a plume, which can reach the high atmosphere after being amplified by the exponential decay of the atmospheric density (Kelley *et al.*, 1985). At greater distances, acoustic waves are primarily generated by the Rayleigh surface waves and lead to ionospheric oscillations. These were first detected in the 1960s following very large quakes (e.g., Yuen *et al.*, 1969; Weaver *et al.*, 1970; Leonard and Barnes, 1965; Davis and Baker, 1965), and are now more commonly

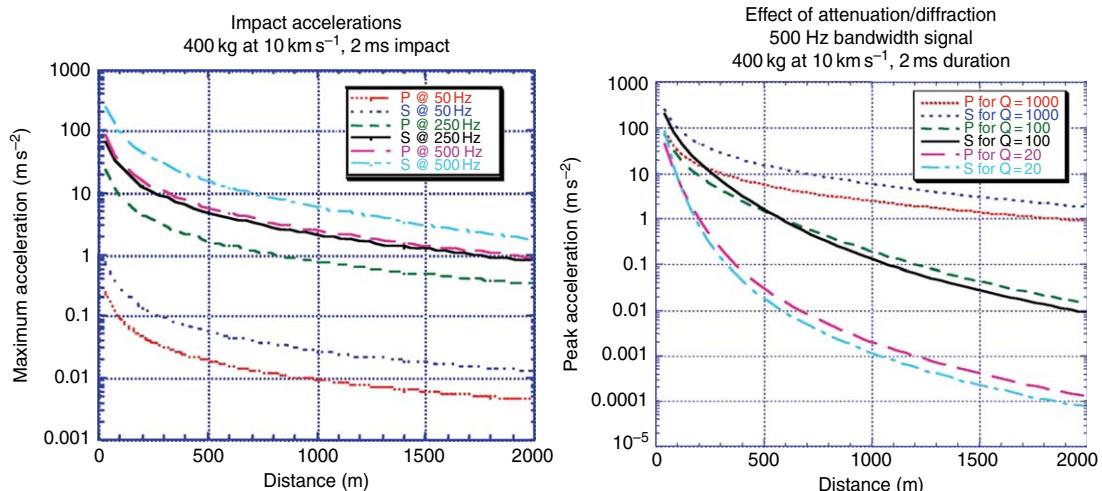


Figure 17 Direct P and S waves amplitude vs distance estimates of the P and S waves for an impact of 400 kg at 10 km s^{-1} for different frequency bands (a) and diffraction/attenuation effects (b). Velocities are 3670 and 2100 m s^{-1} , respectively and densities 2600 kg m^{-3} . A Q of 1000 is assumed in (a). Figure redrawn after Lognonné 2005.

observed for quakes with magnitudes greater than 7–7.5, using GPS (Calais and Minster, 1995; Ducic *et al.*, 2003; Garcia *et al.*, 2005) or Doppler sounders (Artru *et al.*, 2001, 2005).

The excitation of seismic waves by volcanic eruptions (e.g., Kanamori *et al.*, 1984; Zürn and Widmer (1996)) or the continuous excitation of normal modes (Suda *et al.*, 1998; Kobayashi and Nishida, 1998; Tanimoto *et al.*, 1998), called Earth's hum, has been proposed as second examples of such coupling, that time with transfer from the atmosphere to the solid Earth.

In the Hum case, the amplitudes of the signal are small – in the range of >0.3–0.5 nGal between 2 and 7 mHz, but strong enough for identifying the resonance peaks of the fundamental normal modes. Early studies proposed that the turbulences in atmospheric boundary layers (Nishida *et al.*, 2000; Tanimoto, 1999) were strong enough to generate the observed signal. Recent studies show however that the source is located over the oceans (Tanimoto, 2005; Rrie and Romanowicz, 2004, 2006) and that infragravity waves over the continental shelves are much more efficient seismic sources (Webbs, 2007). In summary, the hum on the Earth seems therefore to result from several coupling effects where energy from the atmosphere is first transferred to the oceanic infragravity waves and finally to the Earth's interior. If such a transfer is more efficient than the direct exitation of the normal modes by the atmosphere, what will be the case for a planet without ocean?

This question is addressed to Earth's two nearest neighbors, for which the energy transferred in the atmosphere by the Sun is comparable to the Earth case. The solar flux driving the atmospheric fluid layer circulations is proportional to $(1 - A)/D^2$, where D is the distance to the Sun and A the bound albedo (de Pater and Lissauer, 2001). The relative

ratio of this flux for Venus and Mars compared with the Earth is 0.67 and 0.51, respectively, and demonstrates the need to quantify the efficiency of seismic conversion of this energy for planets without oceans.

As we will see later, all these topics might offer tremendous prospective for new seismological discoveries. For Mars, continuous excitation of free oscillations, if strong enough, could enable the determination of fundamental mode frequencies, even if no large quake is directly detected. For Venus, a remote detection in the atmosphere might provide the opportunity to perform seismic measurements, avoiding the major challenges associated with a long-lived lander able to withstand the 500°C surface temperature. More generally, any planetary body with an atmosphere will exhibit acoustic/seismic coupling between the atmospheric and interior parts of the planet: other bodies of particular interest include Titan and the giant planets, for which no discontinuity exists between the interior and the atmosphere. In this section we explore therefore atmospheric coupling on Mars and Venus, focusing on the expected amplitude of signals. We then provide a short review on the status of giant planet seismology, for which the first seismic observations also remain to be performed.

Let us start by reviewing the main properties related to surface-atmosphere seismic coupling. Density and sound speed for the lower 150 km of the atmospheres of Earth, Venus, and Mars are given in **Table 2** and **Figure 18**. The theory of interior/atmospheric coupling must first take into account the fact that a fraction of the seismic-wave energy is transmitted in the atmosphere when the seismic wave reaches the surface of the solid (or for the Earth, liquid) planet. For a vertically incident acoustic wave, the amplitudes of transmission (t)

Table 2 Atmospheric coupling parameters for Venus, Earth, and Mars

	Crust	Earth	Mars	Venus
Surface sound speed (m s^{-1})	5800	340	214	426
Surface density (kg m^{-3})	2600	1.225	0.0175	65
Acoustic impedance ratio/crust	1	2.76e-5	2.48e-7	1.83e-3
Energy transfer ($E = (Q/\pi)\rho_{\text{air}}c_{\text{air}}/\rho_{\text{int}}c_{\text{int}}$) ($Q = 100$)		8.8e-4	7.9e-7	5.8e-2
High Q atmospheric resonances (mHz)		3.7	2.05	3.10; 4.15

The atmospheric model used for the Earth is the US Reference Atmospheric model. For Mars, it is the mean model from a GCM run of Forget *et al.* (1999).

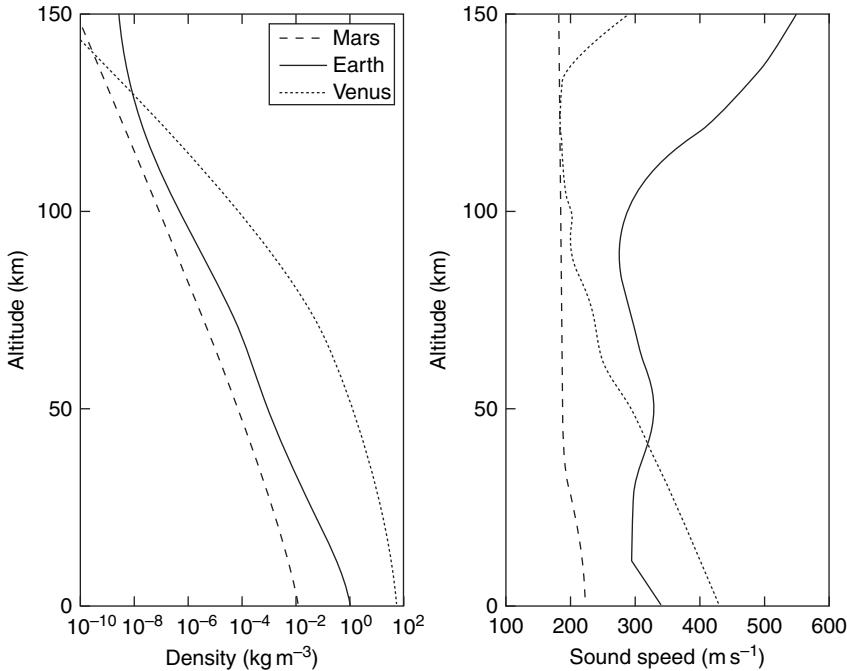


Figure 18 Atmospheric models of the Earth, Mars, and Venus used for the computation of the atmospheric coupling. Note that these models are sensitive to the local time and latitude, as well as to the solar environment. US standard atmospheric model (1976) is used for the Earth, whereas the models of Forget et al. (1999) and Hunten et al. (1983) are used for Mars and Venus, respectively.

and reflection (r) coefficients from the interior to the atmosphere are given by

$$t = \frac{2\rho_{\text{int}}c_{\text{int}}}{\rho_{\text{air}}c_{\text{air}} + \rho_{\text{int}}c_{\text{int}}} \approx 2$$

$$r = \frac{\rho_{\text{int}}c_{\text{int}} - \rho_{\text{air}}c_{\text{air}}}{\rho_{\text{int}}c_{\text{int}} + \rho_{\text{air}}c_{\text{air}}} \approx 1$$
[8]

while the energy transmission (T) and reflection (R) coefficients are

$$T = \frac{\rho_{\text{air}}c_{\text{air}}}{\rho_{\text{int}}c_{\text{int}}} t^2$$

$$R = r^2$$
[9]

and $R + T = 1$. The initial amplitude of the waves in the atmosphere is therefore twice the vertical amplitude of the waves (or the amplitude of the ground, due to continuity), while the relative energy transmitted is 4 times the acoustic impedance ratio. Surface waves of angular order ℓ have a horizontal wavelength of $\lambda = 2\pi a/(\ell + 1/2)$ and will bounce on the surface once per cycle, with amplitude decreasing each time by $e^{-\pi/\mathcal{Q}}$ due to attenuation in the solid part, where \mathcal{Q} is the quality coefficient of the mode. During these successive bounces, the portion of energy transmitted to the atmosphere will

therefore be $E = e(2\mathcal{Q}/\pi)\rho_{\text{air}}c_{\text{air}}/\rho_{\text{int}}c_{\text{int}}$, where e is the partition ratio between the energy in vertical displacement and the total energy, the latter being typically of the order of 0.5. Such energy transmission, given in Table 2, is approximately found in the high ℓ limit of Figure 19.

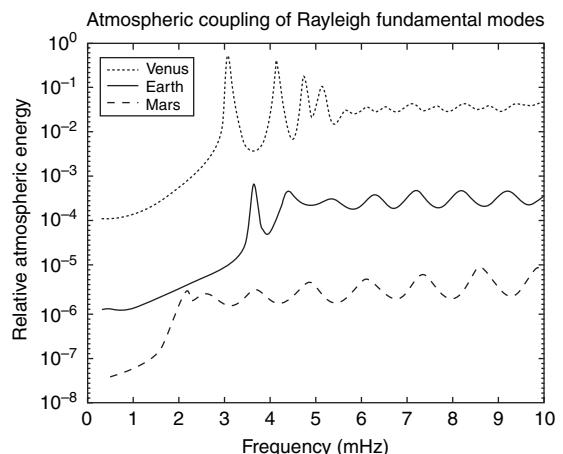


Figure 19 Fraction of the energy of surface waves in the Venus, Earth, and Mars atmospheres for Rayleigh surface waves. Only the first peaks are due to atmospheric resonances. Note that the amplitudes on Mars and Earth are comparable at low frequency (2–3 mHz), due to differences in the atmospheric resonance frequency.

In the atmosphere, in the nonviscous case, the seismic wave equation is given by

$$\begin{aligned} \rho_0 \frac{\partial \mathbf{v}}{\partial t} &= -\nabla p + \rho_1 \mathbf{g}_0 + \rho_0 \mathbf{g}_1 \\ \frac{\partial \rho_1}{\partial t} + \text{div}(\rho_0 \mathbf{v}) &= 0 \\ \text{div}(\mathbf{g}_1) &= -4\pi G \rho_1 \\ \frac{\partial p}{\partial t} &= -\gamma \rho_0 \text{div}(\mathbf{v}) - \rho_0 \mathbf{g}_0 \cdot \mathbf{v} \end{aligned} \quad [10]$$

where ρ_0, ρ_1, p are the equilibrium density, perturbation in density, pressure, and where $\mathbf{v}, \mathbf{g}_0, \mathbf{g}_1$ are the velocity, equilibrium gravity, and perturbation in gravity associated to the mass redistribution. ρ_0 , γ , and G are the equilibrium pressure, the adiabatic index, and the gravity constant. The upward propagation of the waves is modeled by renormalizing the amplitudes in order to account for the exponential decay of the density. Quasi-analytical expressions can be obtained when the atmospheric structure has a vertical scale much larger than the wavelength of the wave and when pure upward propagation is assumed. Gough (1986) shows that under these conditions eqns [10] lead to

$$\left(\frac{\partial^2}{\partial t^2} + \omega_c^2 \right) \frac{\partial^2 \psi}{\partial t^2} - c^2 \frac{\partial^2}{\partial t^2} \nabla^2 \psi = 0 \quad [11]$$

where the cutoff frequency is given in first approximation by $\omega_c = (c/2H_\rho)$ and where $\psi = c^2 \sqrt{\rho_0} \text{div}(\mathbf{v})$ is related to the square root of the acoustic energy $c\gamma_0\rho_0 \text{div}^2(\mathbf{v})$, where $c^2 = (\gamma_0\rho_0)/\rho_0$ is the squared speed of sound. Here, H_ρ is the density height scale given by $H_\rho = -(dL n\rho/dr)^{-1}$. The plane wave solution to eqn [11] obeys the dispersion equation:

$$\omega^2 = \omega_c^2 + k^2 c^2 \quad [12]$$

Waves with frequency lower than the cutoff frequency have negative k^2 and exponentially decaying energy in the atmosphere; waves with frequency higher than the cutoff frequency can propagate upward. This energy is not trapped and the waves progressively lose energy during their upward propagation through viscous and nonlinear processes. Note that other parametrizations are possible, leading to slightly different expressions for the dispersion equation and cutoff frequencies (Beer, 1974; Mosser, 1995). The trapping of free oscillations, which can be used to model the waves with summation techniques, can be modeled when the

full system (i.e., the solid and atmospheric parts of a planet) is considered, and when a radiating boundary condition (Lognonné *et al.*, 1998a) is assumed at the top of the atmosphere. We give here the results of forward modeling based on this theory for the Earth, Mars, Venus, and Jupiter.

For planets with solid parts, the normal modes of the system can be computed for all modes: spheroidal solid modes, atmospheric acoustic modes, and atmospheric gravity modes. The summation of these normal modes then allows the computation of either seismograms for the solid surface and for the atmospheric perturbations, or barograms for the atmosphere as a function of altitude. More details and the theoretical background can be found in Lognonné *et al.* (1998b) and Lognonné and Clévéde (2002). An extension to the viscous case is given by Artru *et al.* (2001). Figure 19 illustrates the efficiency of coupling for Venus, Mars, and the Earth, and shows the fraction of energy of the fundamental Rayleigh modes in the atmosphere when the latter are computed with the atmospheric coupling. Resonance effects appear with a peak in the atmospheric energy; they are observed at frequencies associated with the fundamental and overtones of the atmospheric waves guide and are dependent upon the atmospheric model used.

At these frequencies, a much larger fraction of the seismic waves is transferred to the atmosphere. On the Earth, such resonance frequencies between the atmosphere and interior exist near 3.7 mHz and 4.44 mHz and are preferential windows for studying the atmospheric/interior coupling. Large atmospheric eruptions, such as those of El Chichón in 1982 and Pinatubo in 1991, led to selective excitation of Rayleigh surface waves at these frequencies (Kanamori and Mori, 1992; Widmer and Zurn, 1992; Zürn and Widmer, 1996). The fundamental resonance frequency on Mars is found at about 2.05 mHz while the fundamental and the three first harmonics on Venus are found at 3.10, 4.15, 4.75, and 5.15 mHz. On Venus, the two first overtones are strongly trapped while only the fundamental branch is trapped on the Earth and Mars.

Figure 20 shows the Rayleigh normal modes for angular orders smaller than 50. Below the atmospheric fundamental frequency, amplitudes decay with altitude and full trapping of the modes is observed. At higher frequencies, oscillations appear, as a consequence of the upward propagation associated with energy leakage, and the normal modes also have significant imaginary amplitudes in the

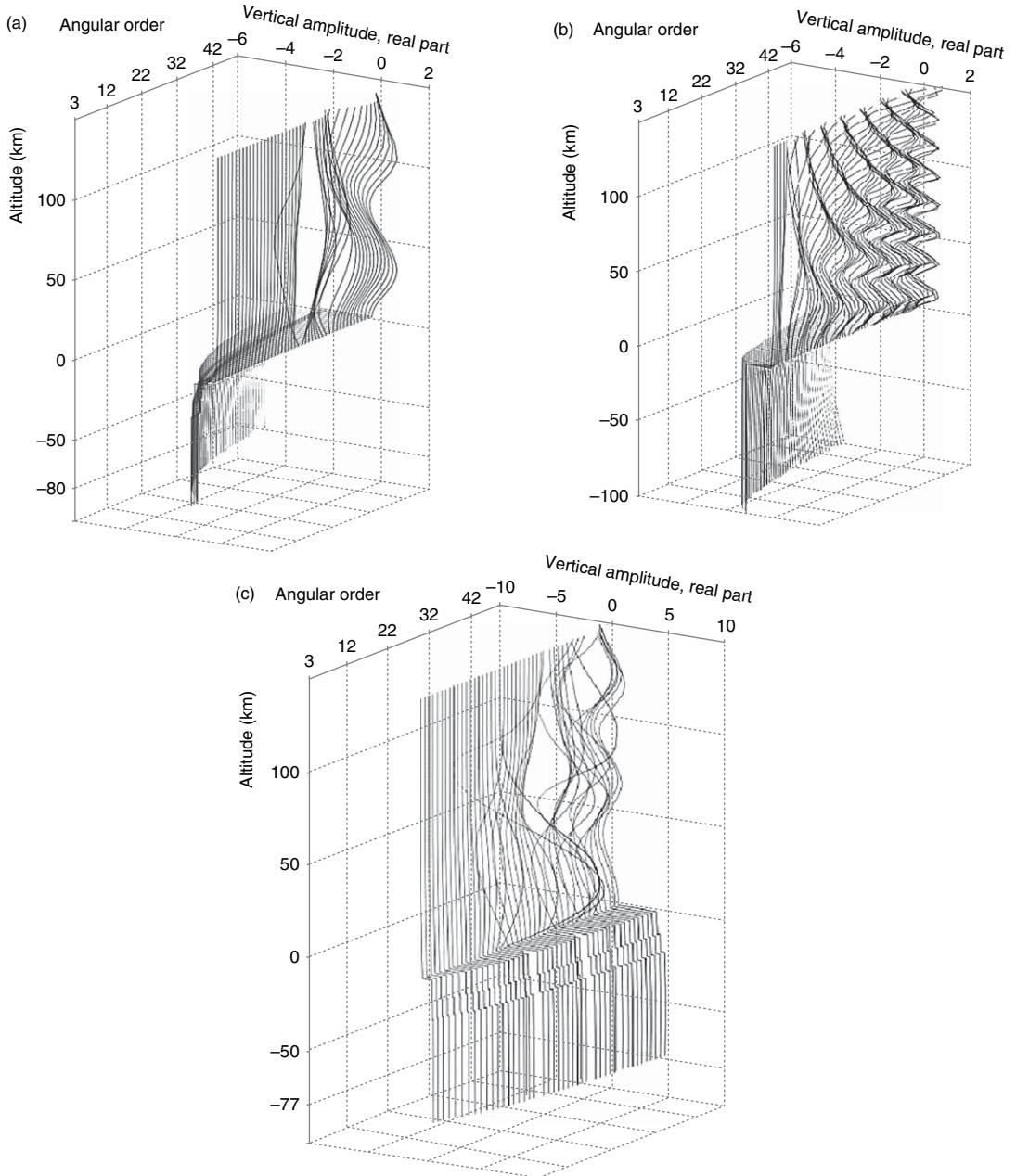


Figure 20 Rayleigh normal modes for the Earth (a), Mars (b), and Venus (c). The amplitudes shown are the vertical displacement, multiplied by the square root of density. Amplitudes are multiplied by 5 for the Earth and by 50 for Mars. For Venus, amplitudes are not amplified. Amplitudes are shown over a vertical scale from 100 km depth to 150 km altitude. Note in all cases the exponentially decaying amplitudes close to the surface for the low angular orders below the cutoff frequency, the single oscillation at the resonance comparable to the amplitude of the atmospheric fundamental modes and the oscillating amplitudes above the resonance frequency.

atmosphere. Due to its smaller size and to a low resonance frequency (associated with the cold atmospheric temperature), this transition for Mars is found for order $\ell = 8$, while it is observed for

$\ell = 28$ and $\ell = 22$ for the Earth and Venus, respectively. The amplitudes are also proportional to the ratio of acoustic impedance between the ground and atmosphere.

10.03.4.2 Mars Hum and Martian Atmospheric Sources

We now address the possibility of continuous excitation of normal modes by the Martian atmosphere. Kobayashi and Nishida (1998) originally proposed amplitudes on Mars to be a factor of 2–3 smaller than on Earth, while amplitudes on Venus were expected to be comparable to Earth. These early studies however considered primarily excitation by turbulence in the boundary layer. See more on the theory and assumptions in Tanimoto (1999, 2001). This theory has been later however criticized by Webb (2007), who shows that an overestimation by several orders of magnitude is done on the normal modes amplitudes.

Can we still imagine that normal modes might be excited with potentially observable amplitudes? Instead of turbulences, let us consider the nonturbulent large-scale winds and atmospheric circulation. As they generate wind Reynolds stresses, they act as a potential seismic source. As shown by Lognonné *et al.* (1996), the density of pressure glut can then be expressed as

$$m_{ij} = (\rho + \kappa \nabla \cdot \mathbf{u}) \delta_{ij} + \rho v_i v_j \quad [13]$$

where ρ is the atmospheric pressure, v_i the i th wind velocity component, ρ the density, κ the bulk modulus of the atmosphere, and \mathbf{u} the displacement field of the normal modes in the atmosphere. In the isotropic case, when wind is neglected, the pressure is much greater than the Hook pressure of normal modes, and we have $m_{ij} = p \delta_{ij}$. The amplitude of a normal mode with angular order ℓ and eigenfrequency ω_ℓ can then be expressed as

$$\begin{aligned} a_\ell^m(t) &= \int_{-\infty}^t d\tau e^{i\omega_\ell(t-\tau)} \iint d\Sigma \bar{Y}_\ell^m(\theta, \phi) \\ &\quad \int_0^{+\infty} dz A_\ell(z) \frac{Mp(z, \theta, \phi, t)}{R\rho_0(z)} \end{aligned} \quad [14]$$

where z, θ, ϕ are the altitude, latitude, and longitude, $a_\ell^m(t)$ is the acceleration of a given normal mode at a time t , $A_\ell(z)$ an excitation term depending on the atmospheric–solid planet coupling, related to divergence of the normal mode eigenfunction times the amplitude of the mode on the recording direction, \bar{Y}_ℓ^m the complex conjugate of spherical harmonic of degree ℓ and order m , p the atmospheric pressure, R the thermodynamical constant, ρ_0 the atmospheric density, and M the mean atmospheric molar mass (Note that $Mp/R\rho_0$ has the dimension of temperature.). Values of the excitation terms, $A_\ell(z)$ at the surface of the planets are shown in Figure 21 for

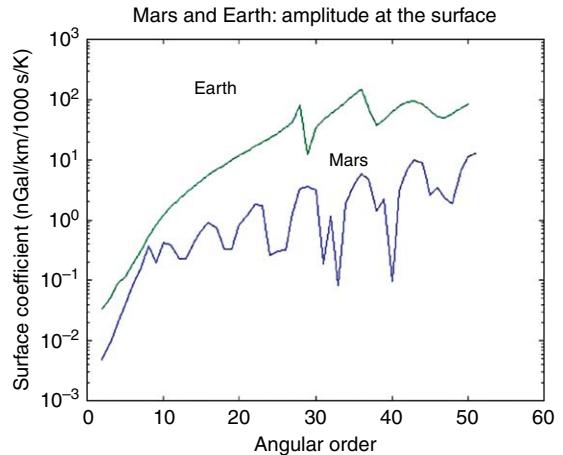


Figure 21 Excitation term $A_\ell(z)$ vs angular order for Earth and Mars (eqn [14]) at the surface of the planet ($z = 0$). Note that at low order, the excitation terms for the two planets are comparable. Due to the high temperature variations, the Martian hum might therefore be more efficient for these low-order terms than on Earth.

the fundamental normal modes in the Earth and Mars cases. On the Earth, up to an angular order of 28–29, acoustic energy is trapped and therefore not radiated. This transition on Mars is achieved at angular order below 10, for all realistic atmospheric models of Mars, leading to a dynamic coupling of the normal modes with the atmospheric circulation of Mars quite efficient for the low-order angular modes. The coupling coefficient is comparable to that on Earth, at angular orders of about 10 and typically a factor 10 smaller for angular orders greater than 20, as a consequence of the smaller atmospheric density. However, temperature fluctuations and winds are significantly larger.

Global circulation models (Forget *et al.*, 1999) can be used to obtain more precise estimates of the continuous excitation of normal modes, through a computation of the pressure glut. Lognonné *et al.* (2001) have performed such study. Figure 22 shows the right-hand side of eqn [14], corresponding to the excitation term related to the pressure field at a given local time and for a classical GCM model just before time integration and multiplication by $e^{i\omega_\ell(t-\tau)}$. Large excitations are observed, including along atmospheric fronts in the South Hemisphere. After integration, such models provide the acceleration amplitudes of the modes (Figure 23). As mentioned above, these estimates are probably a minimum value, because GCMs do not model short-scale circulations in the atmospheric boundary layer. The

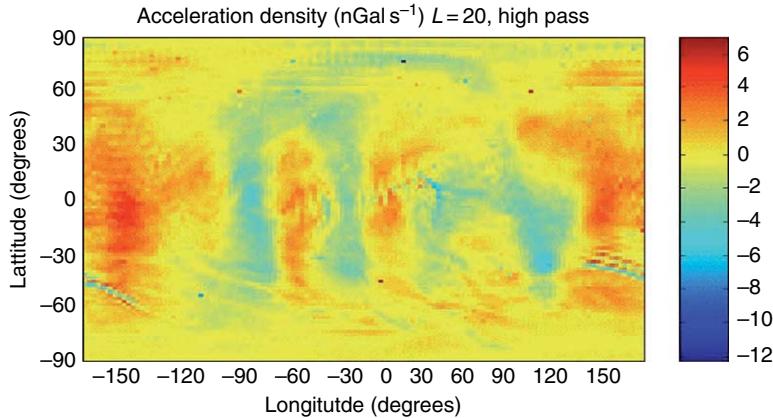


Figure 22 Acceleration field of a Martian GCM model at a given time on the planet, obtained for a 30 days run of the GCM (Forget et al., 1999), corresponding to the term $a_\ell^m(t)$ in eqn [14]. The field is computed for the spheroidal mode ${}_0S_{20}$ and includes integration over altitude, z , but not the Legendre transformation nor the time integration. Note the presence of atmospheric front in the South Hemisphere, which provides highly localized excitation of the normal modes, as well as the large-scale pressure field associated with the daily variation.

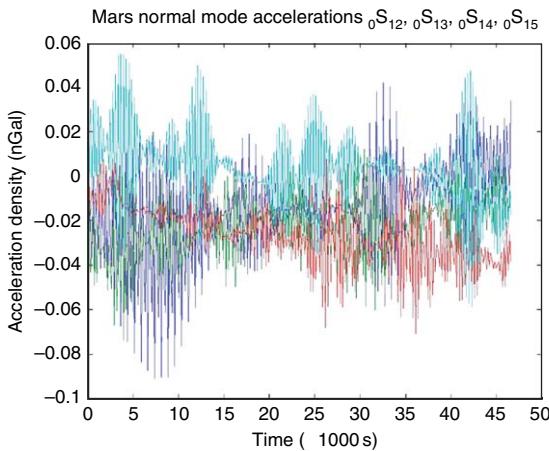


Figure 23 Simulations of the permanent excitation of normal modes, for the spheroidal modes ${}_0S_{12-15}$ on Mars, with a GCM model during 2 weeks. A time interval of about 14 h is shown, illustrating the typical amplitudes variations for these modes, with typical periods from 400 to 300 s. Color code is blue, green, red, and cyan, respectively, for increasing angular orders. Observed amplitudes on the Earth below angular order of 20 are 0.2 nGals.

results show however that a small excitation is achieved by atmospheric global circulation. Moreover, the global circulation is more coherent than on Earth from one day to the next. So if long-duration measurements become available, an increase of the signal by stacking methods can be envisaged. Even though the search for continuous

excitations will be challenging, due to the extremely low amplitudes and the high noise associated with the surface installation of seismometers on Mars, it can open exciting opportunities for the detection of low-angular order normal modes and will benefit from long-lived seismic monitoring, enabling such stacks.

Let us now briefly consider the effect of the atmosphere at shorter periods. The atmospheric wind-induced turbulences are known to be a source of short-period seismic noise in the frequency range 0.1–10 Hz (e.g., Steeples et al., 1997). Recent studies in terrestrial seismology using cross-correlation techniques have shown that the shallow and middle crust can be imaged by extracting the Rayleigh wave Green's Function from long sequences of ambient seismic noise generated by atmospheric sources. The same technique has been used on the Moon at much shorter periods, by using thermal-generated noise, for the investigation of near subsurface (Larose et al., 2005). Such a technique was applied in California on short period (6–20 s) surface waves (Shapiro and Campillo, 2004; Shapiro et al., 2005) and this can also be applied to shorter-wavelength surface waves, sensitive to the uppermost few kilometers of the crust. Martian atmospheric turbulence, especially localized dust devils, can probably generate comparable surface waves. Such a method may be an alternative to future active seismic experiments for studying the Martian subsurface and the associated water reservoirs. This would

obviate the technological and safety challenge of carrying explosive seismic sources to Mars, as was done on the Moon for the Apollo seismic experiments.

10.03.4.3 Venus Atmospheric Seismology

As noted earlier, ionospheric disturbances associated with acoustic waves generated by the Rayleigh surface waves have been observed on Earth in the far field of very large ($M_w > 7$) quakes (e.g., [Ducic et al., 2003](#); [Artru et al., 2004](#); [Garcia et al., 2005](#)) but also in the near-field of smaller quakes. [Calais and Minster \(1995\)](#) reported ionospheric perturbations on GPS data following the Northridge $M_w = 6.7$ earthquake, and observations have been made for magnitudes as low as 5.9, with reported thermospheric perturbations of about 300 K between 300 and 400 km of altitude ([Kelley et al., 1985](#)). For a short review on the ionospheric post-seismic perturbation, see [Lognonné et al. \(2006\)](#). The detection of atmospheric signals associated with quakes might therefore be an interesting alternative to observation performed by seismometers, especially if measurements can be performed by orbiting spacecraft or Earth-based telescopes. However, Venus is the only planet where possibly weaker seismic activity is counterbalanced by a better atmospheric coupling, enabling seismic studies by such an alternative strategy.

We estimate approximately the amplitude of the expected atmospheric signals on Venus associated with large quakes. The amplitudes of seismic waves in the Venusian atmosphere are related first to the acoustic jump at the surface and associated transmission coefficient, and then to the amplification with altitude. At the surface, the pressure is about 90 bars, density about 65 kg m^{-3} , acoustic velocities are slightly higher (425 m s^{-1}) than on Earth and ground coupling (ρc) is about 60 times greater than on Earth. At an altitude of 50 km in Venus' atmosphere, the pressure is comparable to that at the surface of the Earth, but the density is already almost two orders of magnitude lower than that at the Venusian surface. These two effects mean that signals are expected to be about 600 times greater at the same altitude on Venus compared with Earth for the same quake moment magnitude. Good prospects for detection are therefore possible.

The first possibly detectable signal might be the atmospheric and ionospheric perturbation in the vicinity of a Venusian quake. Such signal will consist

of a plume, generating both temperature, possibly brightness-albedo as well as atmospheric/ionospheric oscillations. The characteristics of the acoustic plume generated near the source were first studied by [Garcia et al. \(2005\)](#), who showed that either adiabatic oscillations corresponding to the acoustic waves or nonadiabatic deposition could be generated. For $M_w = 6$, adiabatic temperature oscillations of about 100 K and post-seismic temperature heating of about 10 K are found ([Figure 24](#)). These perturbations, however, take place very near the source, within a distance comparable to the event depth. Their observation will therefore require a global, high-resolution survey, which might be achieved only by spacecraft near Venus. The detection of such signal will be attempted with the Virtis experiment ([Drossart et al., 2004](#)), which is onboard the ESA's Venus Express spacecraft, in orbit around Venus.

In addition to these signals in the vicinity of the quakes, infrasounds and surface waves generated by the quake will also produce atmospheric signals, which can be detected at larger distances. These waves can be computed more precisely following the theory developed by [Lognonné et al. \(1996\)](#), and the difference in the atmospheric coupling can be analyzed by taking, for both the Earth and Venus interior case, the spherical model PREM ([Dziewonski and Anderson, 1981](#)), the latter being corrected with respect to pressure for Venus.

A first and impressive effect of the coupling is the strong perturbation of the quality factor of the fundamental normal modes of the interior ([Figure 25](#)). The presence of the atmosphere is indeed decreasing the attenuation coefficient of Rayleigh modes Q_1 due to the escape of energy. For the Earth, this effect is comparable or smaller than the error in the Q determination. The Earth normal mode ${}_0S_{98}$, for example, close to 10 mHz, has a Q of 115 with the atmosphere and a Q of 118 without the atmosphere, a decrease of only 2.5%. On Venus, we get a Q of 108 for the mode ${}_0S_{93}$ with the atmosphere, and a Q of 121 without, corresponding to a decrease of more than 10%. This means that about 10% of the energy of the Rayleigh waves is dissipated in the Venus atmosphere and confirms the importance of the coupling.

Comparison of synthetic seismograms for the two planets allows further examination of the coupling effects. [Figure 26](#) shows the atmospheric oscillations at 150 km, for the same epicentral distance and seismic source. We find amplitudes about 100 times

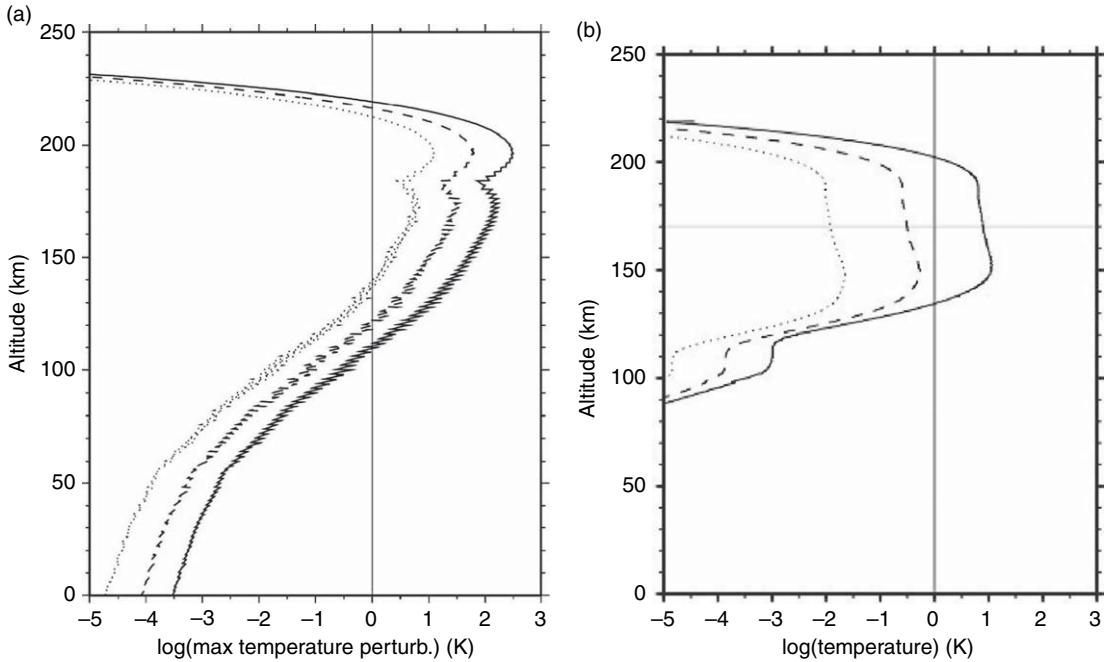


Figure 24 Logarithm of (a) the maximum adiabatic temperature perturbation (in K), as a function of altitude (in km) for different earthquake moment magnitudes: $M_w = 5$ (dotted line), $M_w = 5.5$ (dashed line), and $M_w = 6$ (plain line). Vertical bars are visual. The signals have periods corresponding to the acoustic waves. (b) Same but for the nonadiabatic heating, associated to the acoustic energy deposition in the atmosphere. The temperature anomalies remain for about 4 min, before their atmospheric diffusion. Reprinted from Garcia R, Lognoné P, and Bonnin X (2005). Detecting atmospheric perturbations produced by Venus quakes. *Geophysical Research Letters* 32: L16205.

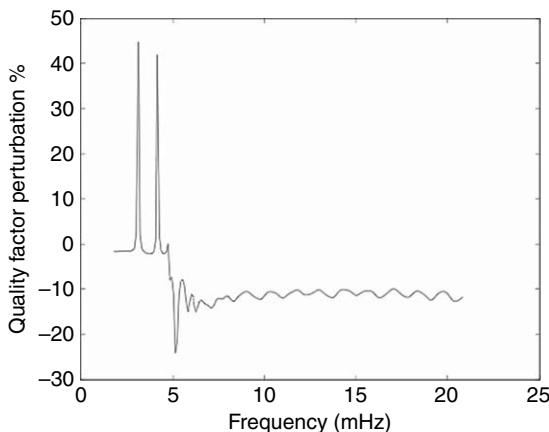


Figure 25 Perturbation of the quality factor for the spheroidal fundamental normal modes. The two peaks are associated with resonances where a significant fraction of energy is transferred in the atmosphere instead of the solid planet, leading to a reduced attenuation. After 5 mHz however, modes loose a significant fraction of their energy in the atmosphere.

larger on Venus than on Earth, reaching 0.3 m s^{-1} at these periods, corresponding to wavelengths larger than 300 km. Amplitudes will be about one order of

magnitude larger at 20 s but are much less, of course, for the low-order angular modes. **Figure 27** shows the amplitude of the latter, for the fundamental Rayleigh mode as well as for the fundamental and two first overtones of the atmospheric waveguide. For a quake releasing a moment of 10^{18} Nm , we get individual mode amplitudes of a few centimeters per second at 150 km for angular orders greater than 30. At 120 km, however amplitudes do not reach 1 mm s^{-1} . The detection of oscillations in the neutral atmosphere (below 120 km) seems therefore not to be feasible, due to the limited length of the Q -cycle of these modes, which is expected to be around a half-day at 5 mHz, for a Q of 200. The existing tools used in Jovian seismology have indeed a typical detection threshold in the range of $0.5\text{--}1 \text{ m s}^{-1}$. Amplitudes at higher altitudes might be however detectable with ionospheric sounders. The Venus ionospheric structure is indeed thinner and the maximum ionization is found at about 150 km. A sounding from the top is therefore possible at such an altitude, contrary to Earth where the maximum of ionization, at about 300–350, prevents the access from the top of the 150–200 altitude layers. With electron density

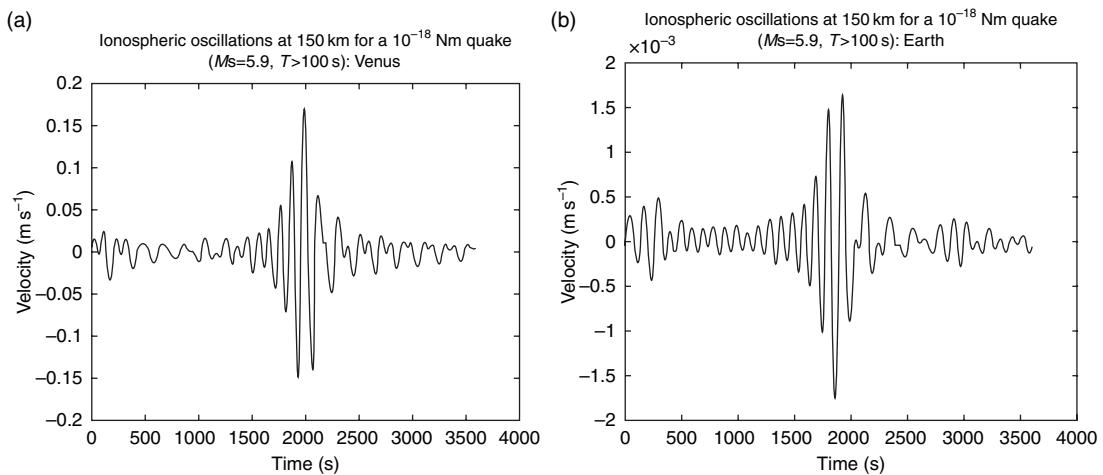


Figure 26 Long-period vertical atmospheric oscillations, for a 10^{18} Nm quake ($M_w = 5.9$) and for period larger than 100 s on Venus (a) and on the Earth (b). Due to the difference in the acoustic coupling at the ground, ionospheric signals at 150 km altitude are about 100 stronger on Venus for the same magnitude. This is about $1.3 M_w$ magnitude. Ionospheric velocity oscillations are about 0.3 m s^{-1} peak-to-peak at these periods, corresponding to wavelengths larger than 300 km. They will be about one order of magnitude larger at 20 s.

ranging from $5 \times 10^3 \text{ cm}^{-3}$ to $15 \times 10^3 \text{ cm}^{-3}$ during the night and from $5 \times 10^5 \text{ cm}^{-3}$ to $5 \times 10^5 \text{ cm}^{-3}$ during the day, this gives electron plasma frequencies ranging from 0.5 to 1.5 MHz during the night and 3–7 MHz during the day. A Doppler sounder, with performances below 1 cm s^{-1} in Doppler measurement (Artru *et al.*, 2004), might offer exciting perspectives for remote-sensing seismology on Venus and has to be considered for a future, high-altitude, Venus orbiter.

10.03.4.4 Giant Planets Seismology

As for the telluric planets, the internal structure of the giant planets is weakly constrained by geophysical data and seismology could provide important new data. For a review on the giant planets' internal structure, see Guillot (2005) and Chapter 10.13.

We review briefly the main seismological properties of giant planets and followed by the state-of-the-art in observations. For a more complete review, including early observations, see Lognonné and Mosser (1993) and Mosser (1996). In contrast to telluric planets, the mantle and possibly even the rocky core are fluid (Guillot, 2005), meaning that only acoustic and gravity waves can be detected. Moreover, the waves are not trapped by a solid-atmospheric discontinuity but only by the effect of the atmospheric density decay. Thus, only waves

with frequencies lower than the atmospheric cutoff frequency are reflected back into the interior.

Figure 28 shows the sound speed and density structure for two typical Jovian interior models, with and without a plasma phase transition (PPT). At this transition, molecular hydrogen is dissociated into metallic or plasma hydrogen (e.g., Saumon and Chabrier, 1989; Saumon *et al.*, 1992). For Jupiter, the cutoff frequency, also shown on **Figure 28**, reaches a maximum of about 4.5 mHz (50 km above the one bar level) and then decreases to about 3.3 mHz, a stable value between 100 and 200 km of altitude. Low-attenuation normal modes can therefore be found only below 3.3 mHz, while leaky modes can exist between 3.3 and 4.5 mHz. No surface waves exist at frequencies higher than the maximum value of the cutoff frequency; a detailed description of the propagation of seismic waves in the Jovian troposphere and stratosphere can be found in Mosser (1995).

The theoretical spectrum of free oscillation of Jupiter was computed for theoretical models of Jupiter in the late 1970s (Vorontsov *et al.*, 1976; Vorontsov and Zharkov, 1981). As for the Earth and other planets, these theoretical studies, as well as more recent ones (Gudkova and Zharkov, 1997, 1999), have shown that the free oscillation frequencies were very sensitive to the internal structure. As noted earlier, free oscillations with frequencies higher than about 3 mHz are not trapped. We show in **Figure 29** the spectrum of free oscillations for the

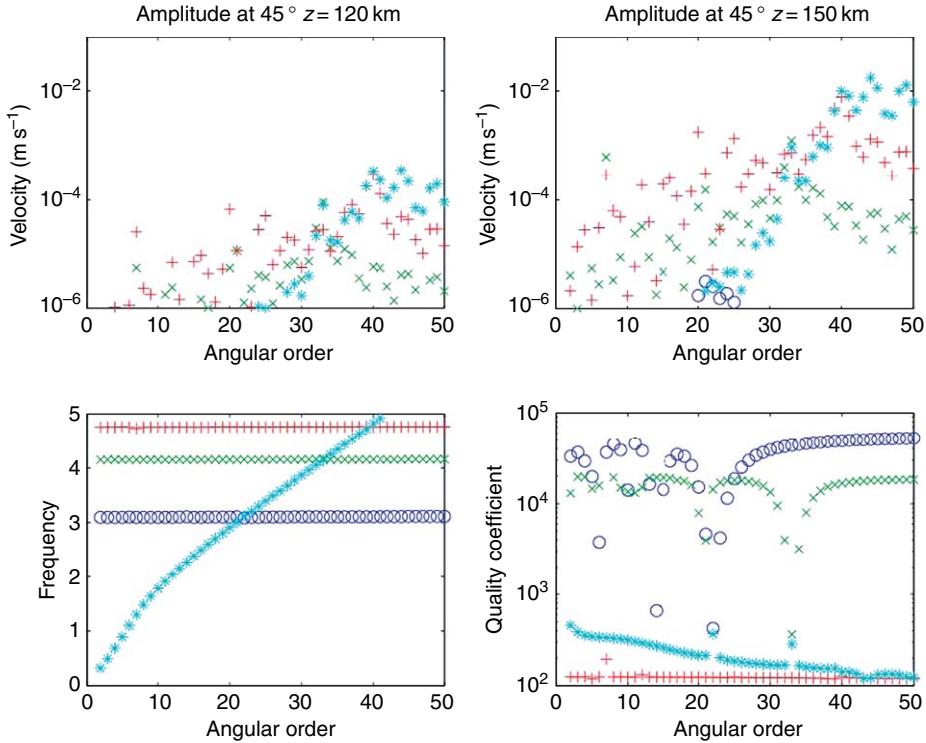


Figure 27 Free oscillations amplitudes (in term of vertical velocities, in m s^{-1}) with respect to angular orders, for a quake with a seismic moment of 10^{18} Nm ($M_w = 5.9$). All these amplitudes are expected to be multiplied by the spherical harmonics values at the location of observations. Sky blue stars are for the Rayleigh fundamental modes. Dark blue circles for the fundamental acoustic modes. Green 'x' and red '+' are for the first and second acoustic modes, respectively. For angular orders below 20, the order of the branches are first the Rayleigh fundamental modes, and then the fundamental, first and second acoustic overtones. In the intraplate cooling hypothesis, such quakes are expected at a rate of two per month (and quakes with moment and therefore amplitudes 10 times larger are expected at a rate of five per year). The top left figure shows the amplitudes of the modes at 120 km of altitude (which can be sounded by optical imaging systems) while the right one shows the amplitudes at 150 km of altitude (which can be achieved by Doppler sounder). The two bottom figures shows the frequencies and quality coefficient of the modes. As no viscosity is taken into account in the model, Q for the fundamental and first acoustic modes are high. The Q of the second overtone is much lower, due to the partial trapping of the modes, while the Q of the Rayleigh modes is also low, due to attenuation in the solid part.

model without PPT shown in Figure 28. Only a bulk Q of 10^6 was considered and the viscosity was neglected. As a result, the low Q values found for frequencies higher than 3 mHz are due mainly to the nontrapping of the waves. Note that the fundamental mode is in fact a gravity mode and that the first acoustic mode is for $n=1$ (Lognonné *et al.*, 1994). Figure 30 shows the amplitudes of the modes, scaled by the square root of the density. Note the trapping of the modes below the cutoff frequency again and the oscillating characters of the acoustic $n=1$ mode for frequencies higher than the cutoff frequency.

The search for the free oscillations of Jupiter started in the late 1980s, with infrared observations

(Deming *et al.*, 1989), Doppler spectrometry observations (Schmieder *et al.*, 1991; Mosser *et al.*, 1991; Cacciani *et al.*, 2001) and Fourier transform interferometry (Mosser *et al.*, 1993, 2000). The main difficulties are

- the limited angular resolution, enabling first only full disk measurement and observation of very low angular orders modes ($\ell \leq 0-3$) and, since 2000, resolution up to $\ell \leq 20-25$;
- the effects associated to the rotation of Jupiter. The normal modes are indeed observed in a non-rotating frame (corresponding to the Earth view for ground observations) and only disk imaging with reprocessing can restore signals associated with a given spherical harmonic Y_ℓ^m ; and

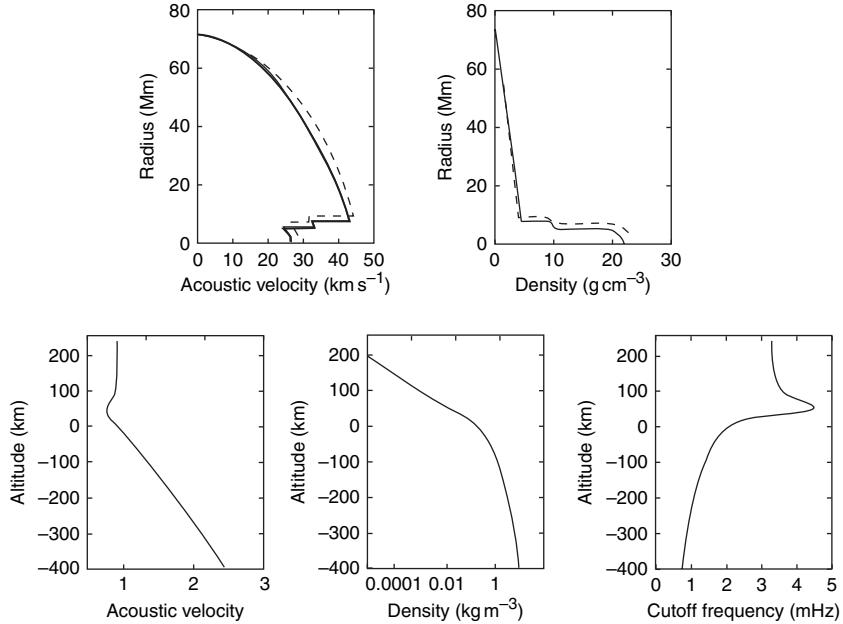


Figure 28 Seismic properties for two different models of Jupiter (Mosser, 1996). The continuous line is a model without plasma phase transition (PPT) discontinuity, while the dashed line is with a PPT. The two upper figures are for the whole Jupiter, while the three bottom ones are for the atmosphere, with the acoustic velocity, the density and the cutoff frequency, respectively.

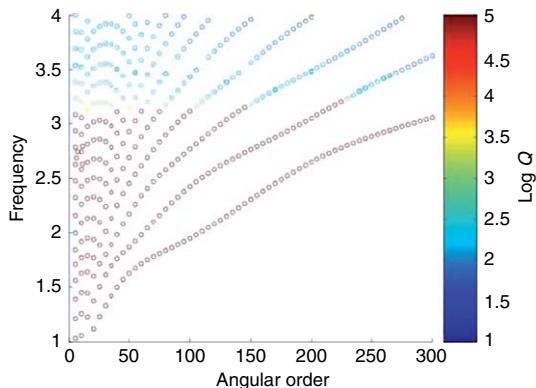


Figure 29 Dispersion curves of the Jovian normal modes, for a model of interior without PPT. Modes are shown with values from 5 to 300, increasing by 5. The color gives the decimal logarithm of the Q . Note that Q strongly decreases for frequencies higher than 3 mHz, as a consequence of the lack of trapping. The tropospheric guided modes, at about 3.50 mHz, are not shown.

- the windowing effects associated with the duration of observations, which has motivated the development of Observatory networks, enabling continuous observation of Jupiter (e.g., Schmider *et al.*, 2003).

The observations performed were not able to detect any free oscillations and an upper limit on

the global oscillation of 1 m s^{-1} has been proposed. However, an excess of signal in the frequency band ($0.5\text{--}1.8 \text{ mHz}$) was detected (Schmider *et al.*, 1991; Mosser *et al.*, 1993). If albedo variations of Jupiter (Cacciani *et al.*, 2001) might be responsible for contamination in the lower part of the spectrum, below 0.7 mHz (Lederer *et al.*, 1995), the upper part might be associated to global excitations. Future observations might be able to reduce the detection threshold. In contrary to telluric planets, global low-frequency oscillations of Jupiter have indeed very high- Q and the phase coherency of the oscillation is practically limited by the excitation processes, as for the Sun, for which oscillations with phase-coherent amplitudes up to 1 year have been observed (e.g., Libbrecht and Woodard, 1991) (p -modes have, however, much shorter phase coherency times of a few days). Continuous future observation might therefore be able to reduce the detection threshold by an order of magnitude after 1 month of continuous observation, putting the limit to about 10 cm s^{-1} . The improvement of performances will also open prospects for the other giant planets, especially Saturn, for which no observations have been reported. Only works on theoretical estimates (e.g., Gudkova *et al.*, 1995) or on the possible impact of the oscillations on the ring

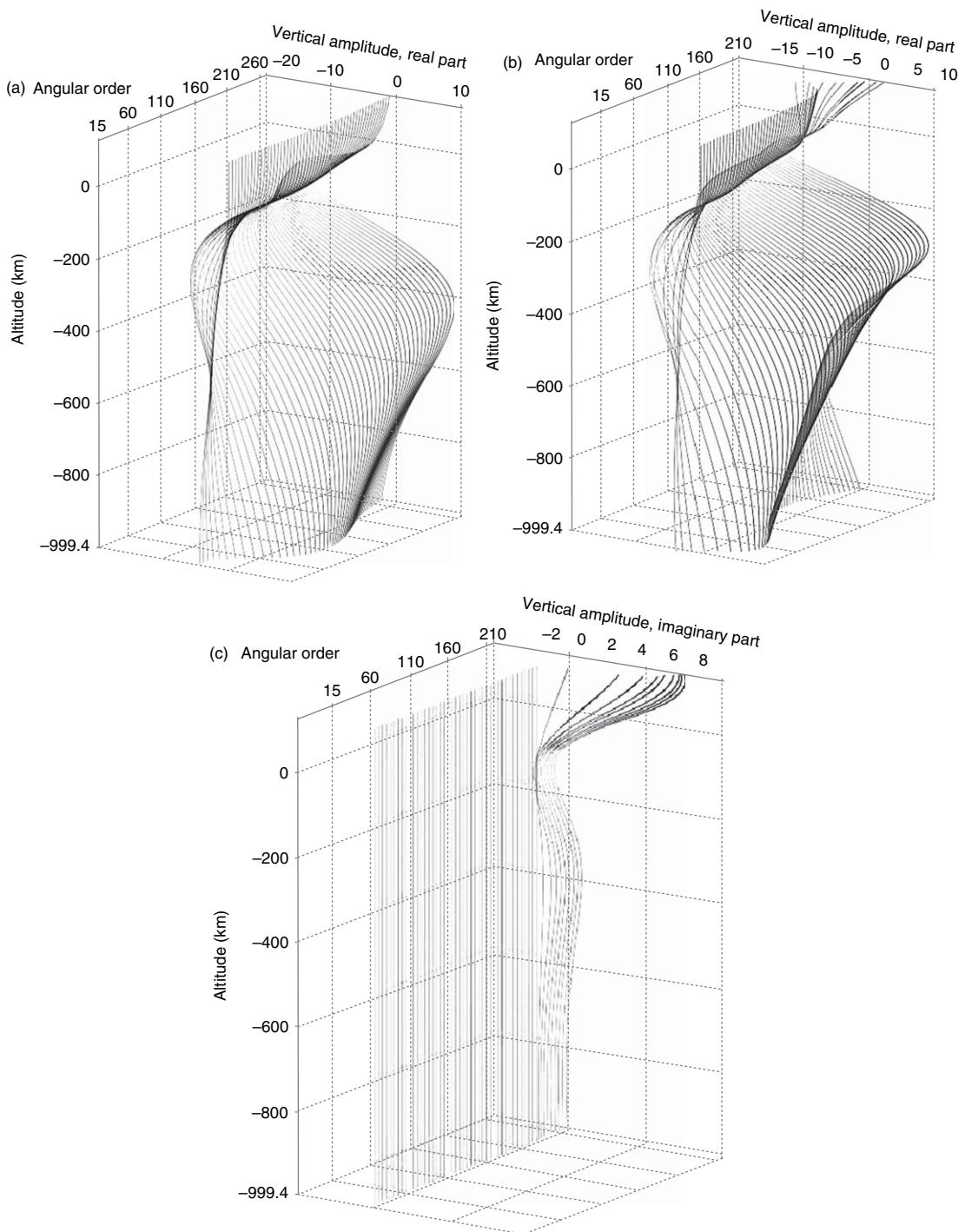


Figure 30 Amplitude of the Jovian normal modes from 1000 km below the 1 bar level to 150 km of altitude. (a) The vertical amplitude of the fundamental mode. All modes are below the cutoff frequency and the imaginary component is almost zero. (b) and (c) The real and imaginary parts of the vertical amplitude of the first overtone, the first seismic modal branch. Note that the energy is focused closed to the surface for the first overtone. Note also the increasing imaginary part, for frequency higher than the cutoff frequency.

structures (Marley and Porco, 1993) have been therefore so far published.

As noted earlier, body waves cannot be trapped in the atmosphere but can in principle be excited by a localized source. The impact of the Shoemaker-Levy 9 comet, due to the high energy of the impact, was such an event. The amplitudes of signals were theoretically predicted for different masses of impactors (Marley, 1994; Kanamori, 1993; Lognonné *et al.*, 1994). For an impact with an energy greater than 10^{21} J, peak-to-peak temperature fluctuations greater than ~ 0.01 K were expected for 10 mHz-frequency P waves, while surface waves below 3 mHz were expected to generate fluctuations in excess of 0.01 K for impacts greater than 2×10^{21} J (Lognonné *et al.*, 1994). No observations were reported by Mosser *et al.* (1996) for impacts A and H nor by Walter *et al.* (1996) for impact R, putting upper limits on the impact energy of $1\text{--}2 \times 10^{21}$ J. Though the impact did not generate detectable body waves, it did produce a ring-like pattern in the Jovian atmosphere, with two rings propagating at 210 and 450 m s^{-1} (Hammel *et al.*, 1995). If such a wave is neither an acoustic body wave nor a surface wave, it is probably a gravity wave, propagating either in the stratosphere (Walterscheid *et al.*, 2000) or in a deeper layer. In the later case, an enhancement by a factor 10 of the water content at the 10 bar depth is necessary for explaining the wave speed (Ingersol *et al.*, 1994; Ingersoll and Kanamori, 1995). See Kanamori (2004) for a detailed discussion of this wave.

10.03.5 The Challenge: Mars Seismology

We now return to the inner solar system, specifically to discuss in detail Martian seismology. In the previous section, we showed that indirect seismological exploration of Mars may be possible (but very challenging and still subject to further theoretical discussions) through studies of continuous excitation of normal modes by the Martian atmosphere global circulation. Here we revisit traditional, quake-based seismology. Though several seismic network missions has been studied in the last 15 years, as described in detail in Lognonné (2005) none of these projects has so far led to launch of a seismic experiment. Such a network would, contribute enormously to our understanding of Mars as a planet, and in our opinion should be one of the highest priorities for imminent exploration of Mars.

The lack of detection of quakes on Mars by the Viking landers (Anderson *et al.*, 1977a, 1977b), along with the failure of the Mars96 mission, including the two small seismometer-equipped autonomous stations and two penetrators, means that no past missions have returned seismic information on the Martian interior. Consequently, the internal structure of Mars is very poorly known. We briefly review the existing constraints on models of the internal structure. We then describe the typical variations of the seismological parameters (e.g., seismic velocities, attenuation), which can be expected from current knowledge of Mars. We finally discuss issues related to the design of a PSE that could provide observations of body waves, surface waves, and normal modes. We focus on considerations relevant to a network with a relatively small number (e.g., 1–5) of seismic stations.

10.03.5.1 Interior Structure of Mars

Seismic and density models of Mars have been proposed by several authors since the Viking mission. The two main types of constraints used in such studies are geophysical and geochemical. Geophysical constraints include the mean density of the planet, obtained from the planetary mass (M) and the planetary radius (a) the mean moment of inertia factor (I) or the moment of inertia with respect to the rotation axis (C), and the k_2 Love number, associated to the gravity signal generated to the Sun tide (see Chapter 10.02 for a review). Okal and Anderson (1978) proposed a model based on the PREM, with a core size adjusted to fit only the mean density. In this model (and similarly for other models) the pressure range in the entire Martian mantle corresponds to that in Earth's upper mantle.

A second set of constraints is provided by geochemical inferences of the bulk composition, based on analyses of the composition of Martian meteorites. One of the most-used models, proposed by Dreibus and Wänke (1985) assumes that the bulk silicate Mars composition is a mixture of C1 material and volatile-depleted C1 material, with a Mars crustal composition constrained by the SNCs. Other model mixtures have been proposed, in order to match the $\delta^{17}\text{O}/\delta^{18}\text{O}$ ratio: a 55% ordinary chondrite H and 45% enstatite chondrite EH (Sanloup *et al.*, 1999), or a 85% H chondrite, 11% CV chondrite, and 4% C1 chondrite (Lodders and Fegley, 1997; Lodders, 2000). One of the important features of these models is the enrichment in iron in the mantle (around twice the Earth's value). As illustrated by Mocquet *et al.* (1996),

the presence and depth of seismic discontinuities in the mantle of Mars will be related to the iron content if the mantle is olivine rich (about 60% in volume). A pyroxene-rich mantle (e.g., Sanloup *et al.*, 1999) should display different discontinuities with smaller amplitudes of the seismic jumps (Verhooven *et al.*, 2005).

Several models have proposed revised estimates of the inertia factor, as compared with that in Sohl and Spohn (1997) and Zharkov and Gudkova (2000), who used a value of $C = 0.365 Ma^2$, corrected for the nonhydrostatic contribution of the Tharsis bulge (Kaula, 1979). This value is close to that obtained later from the measurement of the precession rate of the planet by the Pathfinder mission ($C = (0.3662 \pm 0.0017)Ma^2$; Folkner *et al.*, 1997). More recent estimates that include data from the Mars Global Surveyor (MGS) mission are slightly lower ($C = (0.3650 \pm 0.0012)MR_e^2$, where R_e is the equatorial radius; Yoder *et al.*, 2003). This gives a value of $C = (0.3660 \pm 0.0012)Ma^2$ when the mean planetary radius is used. All these values should be corrected from the gravitational oblateness \mathcal{J}_2 (Munk and MacDonald, 1960) and from the \mathcal{J}_{22} term (Sohl

et al., 2005) in order to get the mean moment of inertia:

$$I = C - \frac{2}{3}\mathcal{J}_2 Ma^2 - \frac{8}{3}\mathcal{J}_{22} Ma^2 \quad [15]$$

These corrections reduce the C value by $1.30e-3$ and $1.68e-4$ respectively (Sohl *et al.*, 2005) leading to $C = (0.3645 \pm 0.0012)Ma^2$ for the Yoder *et al.* (2003) value. The other geophysical datum measured to date is the Love number k_2 , obtained from gravity analysis of orbiting spacecraft. This value is, however, poorly known. If, following Yoder *et al.* (2003) and Bills *et al.* (2005), the value of k_2 is about 0.15 or larger, then only models with large core radius will be possible, as shown by van Hoolst *et al.* (2003). However, lower values for k_2 in the range of 0.10–0.13, have been determined by other authors (Smith *et al.*, 2003; Balmino *et al.*, 2005). In general, as shown in Figure 31, these data support iron cores with some light element(s). Light elements that have been proposed in addition to sulfur (e.g., Dreibus and Wänke, 1985), include hydrogen (Gudkova and Zharkov, 2004), silicon (Stevenson, 2001; Sanloup *et al.*, 2002) and carbon

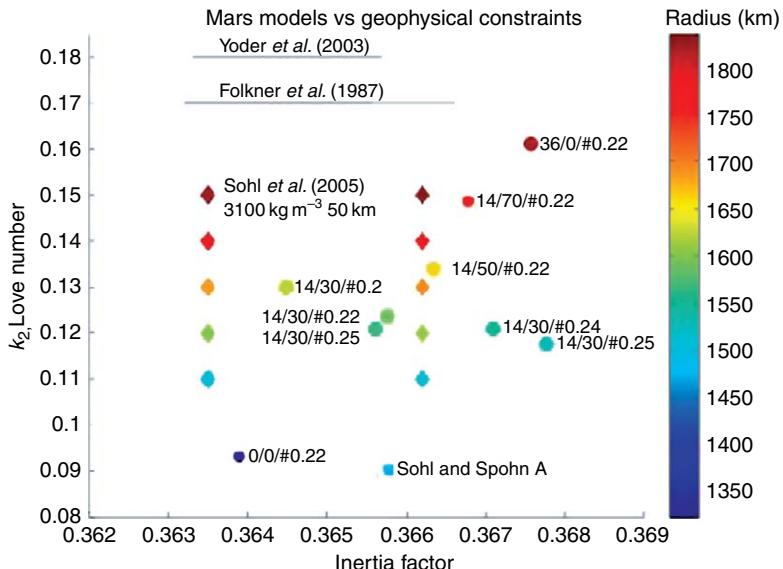


Figure 31 Fit of the computed k_2 and inertia factors for several models of Sohl and Spohn (1997) and Gudkova and Zharkov (2004) (circles) and Sohl *et al.* (2005) (diamonds), the latter for a crust of about 50 km and a crustal density of 3100 kg m^{-3} . Only model A of Sohl and Spohn (1997) is shown, model B with a 0.3566 value for the inertia factor being too far from the observation range. The black horizontal lines are, from right to left, the determinations of I/Ma^2 from Folkner *et al.* (1997) and Yoder *et al.* (2003) after correction for the extrapolated value of I/Ma^2 given by Sohl *et al.* (2005). The core radius is given by the color scale and note that a change of about 10 km is able, for the models taken from Sohl *et al.* (2005) to change the moment of inertia from 0.3635 to 0.3662. For the Gudkova models, the amount of sulfur, hydrogen, and #Fe are given. The two first are in weight % and molar %, respectively, while the third is the molar Fe ratio, defined as $\text{Fe\#} = \text{Fe}^{2+}/(\text{Mg} + \text{Fe}^{2+}) \times 100$.

(Kuramoto, 1997), while oxygen, another light element possibly present in the Earth's core is less likely on Mars (Rubie *et al.*, 2004). A relatively low density core with high H content, as suggested by Gudkova and Zharkov (2004), seems less supported by the new reduced values of the moment of inertia. A detailed analysis of the models proposed by these authors shows that the core's size variations are about ± 150 km (Figure 31). This is large enough for major differences in term of planetary evolution. As an example, the presence of a perovskite-bearing lower mantle becomes impossible, due to insufficient pressures in the mantle. An endothermic phase transition from spinel to perovskite has been suggested to be required in mantle convective models for the early formation of Tharsis (Harder and Christensen, 1996; van Thienen *et al.*, 2006). The size and state of the core are also critical to understanding the cessation of an early Martian dynamo (Acuña *et al.*, 1998; Stevenson, 2001; Breuer and Spohn, 2006; see Chapter 10.09). Thus, the determination of the core size is a key objective for seismology, not just for the knowledge of present-day interior structure, but in order to constrain models for the geodynamical evolution of Mars.

Two other deep interior parameters that affect seismic-wave propagation and therefore the design of future seismic experiments on Mars are the physical state of the core (liquid or solid) and mantle attenuation. A liquid (outer) core may result in a seismic shadow zone, as observed on the Earth, and a highly attenuating mantle (due to volatiles, and perhaps locally warm temperatures) can strongly reduce the amplitudes of the short-period body waves, as may be the case on the Moon (Nakamura, 2005). A liquid core is likely and is supported by Phobos tidal acceleration value (Bills *et al.*, 2005) and by the large k_2 value. As noted by Lognonné and Mosser (1993) and by Zharkov and Gudkova (1997), the low value of the Martian Q at the Phobos tidal period would imply an unrealistically low intrinsic Q of the planet if the core is solid. In contrast, a Q corresponding to silicate material is found for a liquid core (Lognonné and Mosser, 1993). The extrapolation from the tidal Q to the seismic Q can be performed using the frequency dependence proposed by Anderson and Given (1982). Assuming that Phobos' frequency is in the seismic bandwidth, we obtain a shear Q of about 200 at 100 s (using a power law of 0.15; Lognonné and Mosser, 1993; Zharkov and Gudkova, 1996), as compared to a Q of 85 at the 5 h 32 min Phobos

period. This is between the lower-mantle PREM (Dziewonski and Anderson, 1981), shear Q (about 310), and the upper mantle one (about 140). Moreover, we can assume the higher Q might be found in the upper mantle, depending on the actual temperature profile.

Other details of mantle structure remain largely unknown. A key parameter, in terms of seismological inferences, will be the iron content of the mantle. Mocquet *et al.* (1996) have studied the effect of increasing iron content from 10% to 40% and many other models have Fe ranging from 20% to 30% (Dreibus and Wänke, 1985; Sanloup *et al.*, 1999; Gudkova and Zharkov, 2004). The main effect of an increase in iron is to smooth the seismic discontinuities associated with α -olivine to β -spinel and β -spinel to γ -spinel, as seen in Figure 32 for both the density and acoustic velocity v_k ($v_k^2 = v_p^2 - 4/3 v_s^2$). For an iron content of 20%, the width of the transition zone is seen to be between 50 and 100 km. This is in contrast to the Earth, for which both the reduced iron content and the higher pressure gradient result in sharp discontinuities, with a typical width of less than 10 km (e.g., Tibi and Wiens, 2005). Reflections from mantle discontinuities on Mars might be expected to be less impulsive than on Earth and observable only at long periods, when the seismic wavelength is much greater than the width of the discontinuity, that is, below frequencies of 0.05–0.1 Hz. Consequently, these phases might be observed only for the largest quakes. In summary, we can expect that, as for the Moon, most of the mantle information will be obtained from direct P and S waves and from phases reflected at the core and the surface (PKP, Pcp, ScS, PP, etc.).

10.03.5.2 Martian Seismic Noise

One of the main reasons for the lack of seismic activity detected by the Viking seismic experiment was the wind sensitivity of the seismometer. This wind noise was primarily the consequence of the seismometer location. The seismometer was deployed on the platform of the lander; coupling with the ground was provided by the elastic legs of the lander and the seismometer detected the vibrations of the lander due to noise. We can expect better success from a future mission with installation of the seismometer directly on the ground. A rather simple wind-shield covering the seismometer can strongly reduce wind effects. Field tests have demonstrated performance on the Earth at a noise level almost comparable to the Low Noise model at periods of a

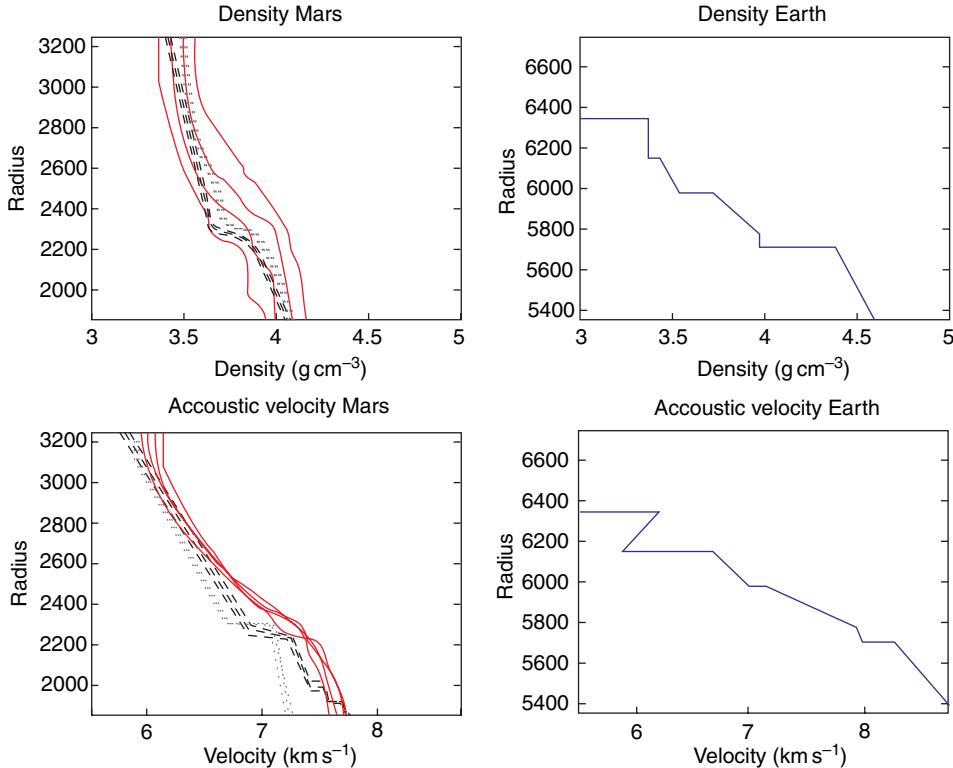


Figure 32 Model M1 to M4 of Gudkova and Zharkov (2004) with Fe contents from 0.20 to 0.25 (0.20–0.22–0.24–0.25) (dash-dot lines), models of Mocquet *et al.* (1996) with Fe content from 10% to 40% (0.1–0.2–0.3–0.4) (solid line) and models of Verhoeven *et al.* (2005) with Fe contents from 0.09 to 0.28 (0.09–0.2–0.25–0.28) (dotted line). The difference in depth of the discontinuities is mainly related to differences in the core–mantle boundary temperature. A temperature of approximately 2100 K is taken at the core–mantle boundary for the model of Gudkova and Zharkov (2004), whereas a more complex model, with temperature inversion in the mantle and mantle temperature approximately 500 K colder at the depth of the discontinuities, is taken by Mocquet *et al.* (1996). These colder temperatures shift the discontinuity to a shallower depth. The temperatures for models of Verhoeven *et al.* (2005) are those of Breuer and Spohn (2003).

few tens of seconds (Lognonné *et al.*, 1996) and we can therefore expect noise levels as low as $10^{-9} \text{ m s}^{-2} \text{ Hz}^{-1/2}$ in the bandwidth 1–0.05 Hz, dominated on Earth by the microseismic peaks.

As Mars has neither an ocean (the major source of terrestrial noise between 0.07 and 0.14 Hz) nor human activity, we can expect low seismic noise compared with Earth. Wind-generated noise has been estimated by Lognonné and Mosser (1993), and temperature-generated noise by van Hoolst *et al.* (2003). Ground accelerations produced by the direct deformation of the Martian surface by the wind pressure fluctuation could have peak-to-peak amplitudes on the order of 10^{-9} m s^{-2} in the range of 0.1–0.01 Hz, during relatively calm conditions with wind speeds of $\sim 4 \text{ m s}^{-1}$ (Lognonné and Mosser, 1993). Moreover, Lognonné *et al.* (1996) have shown that even on Earth, prototype Martian seismometers can reach such noise levels at a quiet site

when protected by a light windshield. In contrast to the Viking mission, temperature variations may be the major source of noise for surface-installed VBB (very broad band) seismometers. However, as shown by van Hoolst *et al.* (2003), most of the temperature variations on Mars are associated with the daily cycle and can be approximated by a Fourier series with the fundamental mode (of 24 h 56 min) and its harmonics. Moreover, thermal insulation of space-qualified seismometers can reach time constants larger than 2 h (D. Mimoun, personnal communication). This will reduce temperature variations experienced by the sensor by a factor of ~ 7 at 1 mHz, and ~ 700 at 0.1 Hz, leading to temperature fluctuations smaller than a fraction of a degree in the seismic band. An installation with a noise level comparable to the terrestrial low noise model (e.g., about $10^{-9} \text{ m s}^{-2} \text{ Hz}^{-1/2}$ in the seismic bandwidth, see Peterson, 1993) is therefore a realistic challenge,

and could be further improved with an environmental correction performed using pressure and temperature data (e.g., Beauduin *et al.*, 1996).

10.03.5.3 Body-Wave Detection

Due to the lower event magnitudes as compared with Earth (see **Figure 13** in Section 10.03.3), the best signal-to-noise data will probably be found in the bandwidth of body waves and regional surface waves. The main limitations for body-wave detection will be attenuation and diffraction. The importance of attenuation on Mars was originally pointed out by Goins and Lazarewicz (1979), who showed that the Viking seismometer, operating with a free frequency of 4 Hz, was unable to detect remote events due to attenuation. For a Q of 325, obtained as explained in Section 10.03.5.1, and a velocity of 5 km s^{-1} , amplitudes of S waves are reduced by a factor of 10 after 1200 km of propagation at 1 Hz. Amplitudes at 4 Hz are reduced by the same amount only after 300 km of propagation. The amplitudes of body waves are plotted in more detail in **Figures 33(a) and 33(b)** after Mocquet (1998). The amplitudes are computed for two frequency bands (0.1–1 Hz, 0.5–2.5 Hz), for an isotropic surface source (quake) with a seismic moment of 10^{15} Nm . Transmission across a crust–mantle boundary is taken into account, as well as geometrical spreading and attenuation, for a Mars model compatible with the present *a priori* knowledge and a shear Q provided by $Q_\mu = 325 - \log_{10} T$, where T is the period.

At frequencies higher than 0.5 Hz (0.5–2.5 Hz), the amplitudes of the body waves decrease rapidly with epicentral distance (**Figure 33**). Amplitude is relatively independent of epicentral distance only for P waves in the frequency band 0.1–1 Hz. Consequently, this frequency band, with a 1 ng resolution, was chosen for the OPTIMISM seismometer (Lognonné *et al.*, 1998a). At higher frequencies (0.5–2.5 Hz bandwidth); however, amplitudes decrease rapidly with source–receiver distance and are less than S-wave amplitudes at the longest periods (**Figure 33(a)**). For an instrument with noise characteristics of $1 \text{ ng Hz}^{1/2}$, noise levels of $\sim 2.5 \text{ ng rms}$ will be expected in the 0.5–2.5 Hz frequency bandwidth, limiting the detection of S waves at large epicenter distance to quakes with a moment larger than 10^{15} Nm .

Short-period body waves are also sensitive to scattering. This has been demonstrated on the Moon and is due to the intense fracturing of the crust, from meteoritic impacts as well as tectonic structure (e.g., Dainty and Toksöz, 1977; Nakamura, 1977). On the Earth, scattering is also very strong in volcanic regions. These combined effects suggest that significant scattering may occur in volcanic areas on Mars, particularly in the Tharsis region. For body waves, scattering reduces the peak-to-peak amplitudes by producing conversions (mainly P to SV, P to SH), and spreads this energy in time. This leads to energy transfer from P to S waves up to an equipartition given by $(E_p = v_s^3/v_p^3)(E_s/2)$, where E_p, E_s are the energy of P and S waves and v_p, v_s are the velocities of P and S waves, respectively (e.g., Aki, 1992; Papanicolaou *et al.*, 1996).

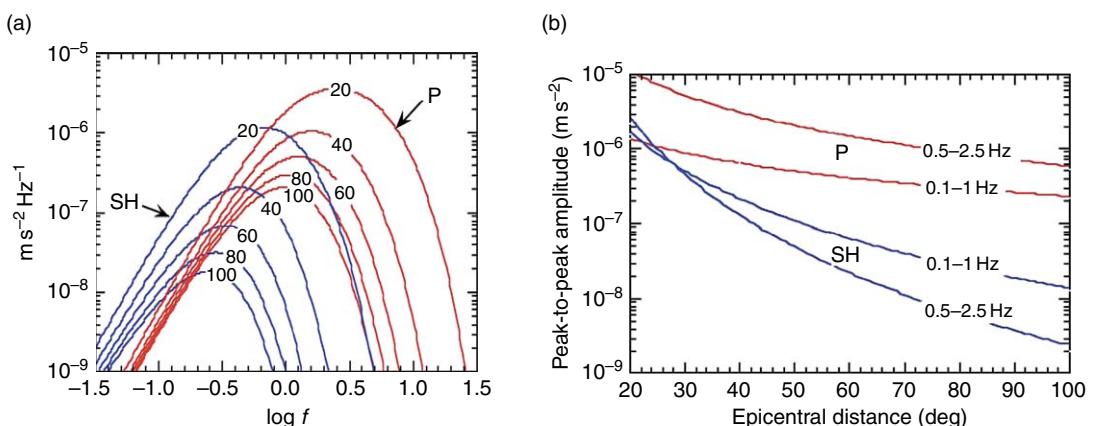


Figure 33 (a) Fourier transformed amplitude of the synthetic P and S body-wave packets. Amplitude is plotted with respect to frequency for different epicentral distances, in degree. (b) Maximum peak-to-peak amplitude in the frequency band for various epicentral distances for a 10^{15} Nm seismic moment quake at the surface. For a source below the surface, the amplitudes are divided by 2, assuming both P and pP arrivals. Reproduced from Mocquet (1998) A search for the minimum number of stations needed for seismic networking on Mars. *Planetary and Space Science* 47: 397–409, with permission from Elsevier.

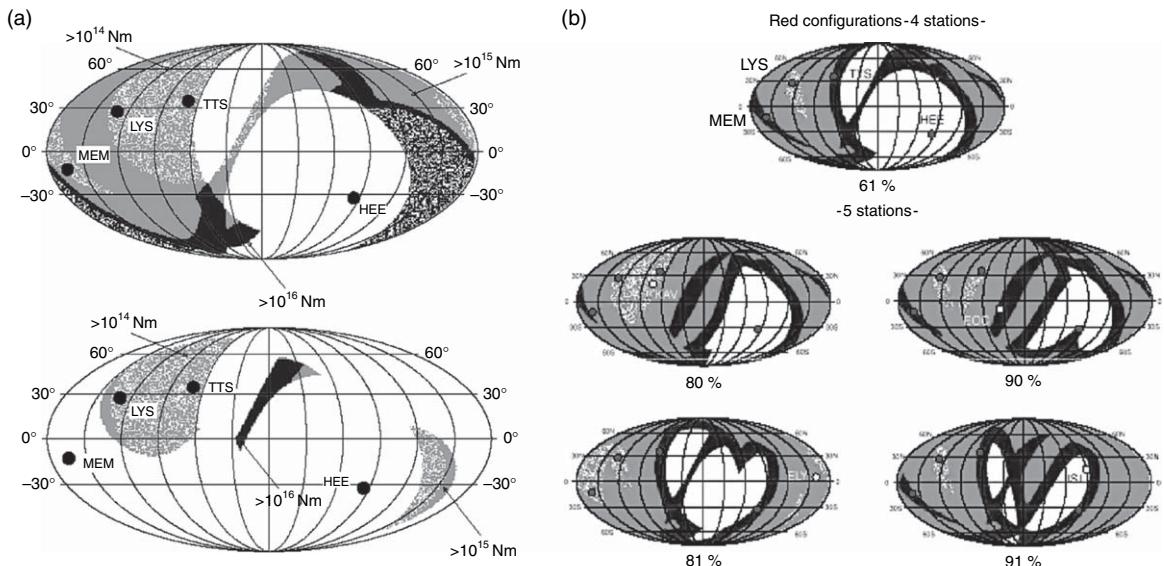


Figure 34 (a) (Top) Detection zone of P and S waves for one of the studied configuration in the former Netlander project (MEM, LYS, TTS, and HEE are, respectively, Menomnia Fosae, Lycus Sulci, Tempe Terra South, and Hellas East). The white zone on to the east of TTS corresponds to the joint shadow zone of MEM and HEE. Note that the detection efficiency is very high in the Tharsis area, where small quakes of seismic moment of 10^{14} Nm might be detected. (Bottom) Detection zone for a detection of a PKP wave in at least one station and of P and S waves in the three other stations. The success rate is 16%, leading to about 22 quakes detected during the two years of operation. Reproduced from Lognonné P, Giardini D, Banerdt B, et al. (2000). The NetLander very broad band seismometer. *Planetary and Space Science* 48: 1289–1302, with permission from Elsevier. (b) Zone covered by a successful detection of P and S waves on at least three stations for a network with four or five stations. The top figure recalls the detection zones with only four stations, while the other are for five stations. In this case, the red dots are the 4 nominal landing sites of the Netlander mission. The white dot is a possible new landing site. The shaded zone corresponds (from gray to black) to detection threshold from magnitude 3.5 to 5. Note the large increase of the total surface covered by the network given below each figure. Note also that the dashed gray zone, corresponding to the detection of small quakes of seismic moment of 10^{14} Nm, has either a larger surface or appears in several different spots, each of them being associated with a successive detection on the three closer stations.

This scattering mainly affects P waves. For shallow quakes it will reduce the amplitude of the P waves near the source and the receiver: first during the downward propagation from the crust to the mantle, near the source, and again during the upward propagation from the mantle to the surface, near the receiver. The effect can be a reduction by a factor of 10 of the P-wave energy, with larger amplitude reduction due to the length of the coda.

The potential efficiency of a future network configuration can be investigated using these results, as shown by Mocquet (1998). **Figure 34(a)** details such a study for a possible network configuration, assuming total detection for signals with accelerations greater than 10^{-8} m s^{-2} peak-to-peak. They show that a rate of 60% can be achieved for the detection of quakes with seismic moment greater than 10^{14} Nm, that is, corresponding to Earth moment magnitude greater than 3.2. As shown on **Figure 13**, depending on the seismic activity, between 40 and 400 quakes with at

least this seismic moment are expected to occur during one Martian year. Note that the four station network configuration, in contrary to the lunar seismic network, will allow a much better characterization of the core, with antipodal stations near Hellas Planitia. **Figure 34(b)** shows the increase in coverage for a network configuration with an additional station, which would also provide redundancy. Source localization errors will affect the determination of the seismic velocities and might be large for network with only a few stations. These have been discussed by Gagnepain-Beyneix et al. (2006), for the lunar case.

10.03.5.4 Normal Mode Excitation and Tidal Observations

While the extraction of internal structure information from the direct body waves will need precise event localization, and hence at least three stations for traveltimes analysis (six traveltimes data enabling the

determination of the four source parameters and two mean mantle velocities), surface-wave and normal-mode analyses have the potential to provide useful information on the interior with fewer stations. As noted above, a cumulative activity of 10^{18} – 10^{19} Nm per year means it may be possible to record quakes with moments as large as 10^{18} Nm if the station is operated for one Martian year or more. The excitation of normal modes by Marsquakes was studied by Lognonné *et al.* (1996) and later by Gudkova and Zharkov (2004). Both studies concluded that the observation of normal modes between 5 and 20 mHz with a noise level of 10^{-9} ms⁻² Hz^{-1/2} will be possible from stacked records of quakes with a cumulative 10^{18} Nm moment or from single-record analysis of the greatest quakes (Figure 35). Normal modes will peak, with a signal to noise ratio of about 10, within Q -cycle windows for such quakes, which correspond to moment magnitudes of 5.5 (Lognonné *et al.*, 1996). Due to the size of the planet, however, the frequencies of the modes, for a given angular order, are typically twice those on Earth (Figure 36). The observation of normal modes below 5 mHz, which are the most sensitive to the core size, will be more challenging with such quakes: modes like ${}_0S_2$ might remain below the instrument or station noise, especially because of temperature fluctuations due to operation of the first Mars seismometers at the surface.

As noted by Lognonné and Mosser (1993) and van Hoolst *et al.* (2003), an alternative core structure constraint will be a measure of the solid tide amplitude, and therefore of the associated Love numbers, by a surface sensor. A successful measurement will be mainly determined by the temperature noise and will require a stack of year-long time series. Van Hoolst *et al.* (2003) have shown that a noise level of about 0.5 K rms in a frequency bandwidth of 1 mHz around the phobos tide is expected, based on the Viking temperature data. Both temperature decorrelation and thermal protection can reduce this signal by a factor of 100–250 after data processing, leading to a final noise level of 2.5–1 ng rms after one Martian year of stacking and an error in the core determination equal to 60 km for 1 ng rms of noise.

10.03.5.5 Surface Waves

Though no surface waves were recorded on the Moon by the Apollo seismometers, the better performance at long period expected of the seismometers bound for Mars and the larger magnitude of Marsquakes suggest the possibility of such detections on Mars. By sampling

the crust, lithosphere, and upper mantle, surface waves will be an important source of information. Their group velocities for example are very sensitive to the crustal thickness (Figure 37), with 10% typical variations for crustal variations of 20 km (Figure 10). The sensitivity to the upper mantle will also be important, and the group velocity of surface waves (or the differential group velocity between the fundamental and the overtones) varies by 5–10% for the models of Mocquet *et al.* (1996) with different iron content (Figure 10).

The practical difficulties in data analyses will be related to quake localization errors and to the effect of lateral variations in the surface waves. These errors will be the most critical if only a limited number of stations are deployed and may produce errors in the determination of spherically symmetric models of the mantle. In the band 5–50 mHz, surface-wave analysis could be performed by two stations equipped with three axis seismometers, enabling the determination of the back-azimuth of surface waves. Figure 38 shows that errors in the relative epicentral distance of less than 2.5% can be achieved over a wide equatorial band of the planet ($\pm 20^\circ$ latitude), for two typical landing sites around the Tharsis Bulge, when back-azimuth errors of 15° are assumed. Such measurements could resolve the major differences among currently proposed mantle models (Figure 32). However, limitations due to the sensitivity of surface wave azimuth to lateral variation in crustal thickness means that it will not be possible to detect smaller differences, such as those related to the FeO content in the mantle (Mocquet *et al.*, 1996). On the Earth, such azimuthal variations can reach typically 10° and it is possible that lateral variations in crustal structure on Mars will lead to larger effects (Larmat *et al.*, 2006). However, as observed on the Earth, much of the lateral variations of surface waves can be modeled with *a priori* on the lithosphere and upper mantle (Nataf and Richard, 1996). With the recent improvements in crustal thickness estimates (Zuber, 2001), lithospheric thickness (Belleguic *et al.*, 2005) and crustal heating, we can expect future developments in the modeling of the effect of the crustal and lithospheric lateral variations on surface waves and associated corrections.

10.03.6 Concluding Remarks

Planetary seismology began almost concurrently with planetary exploration, with seismometers onboard the Ranger probes in the early 1960s. The

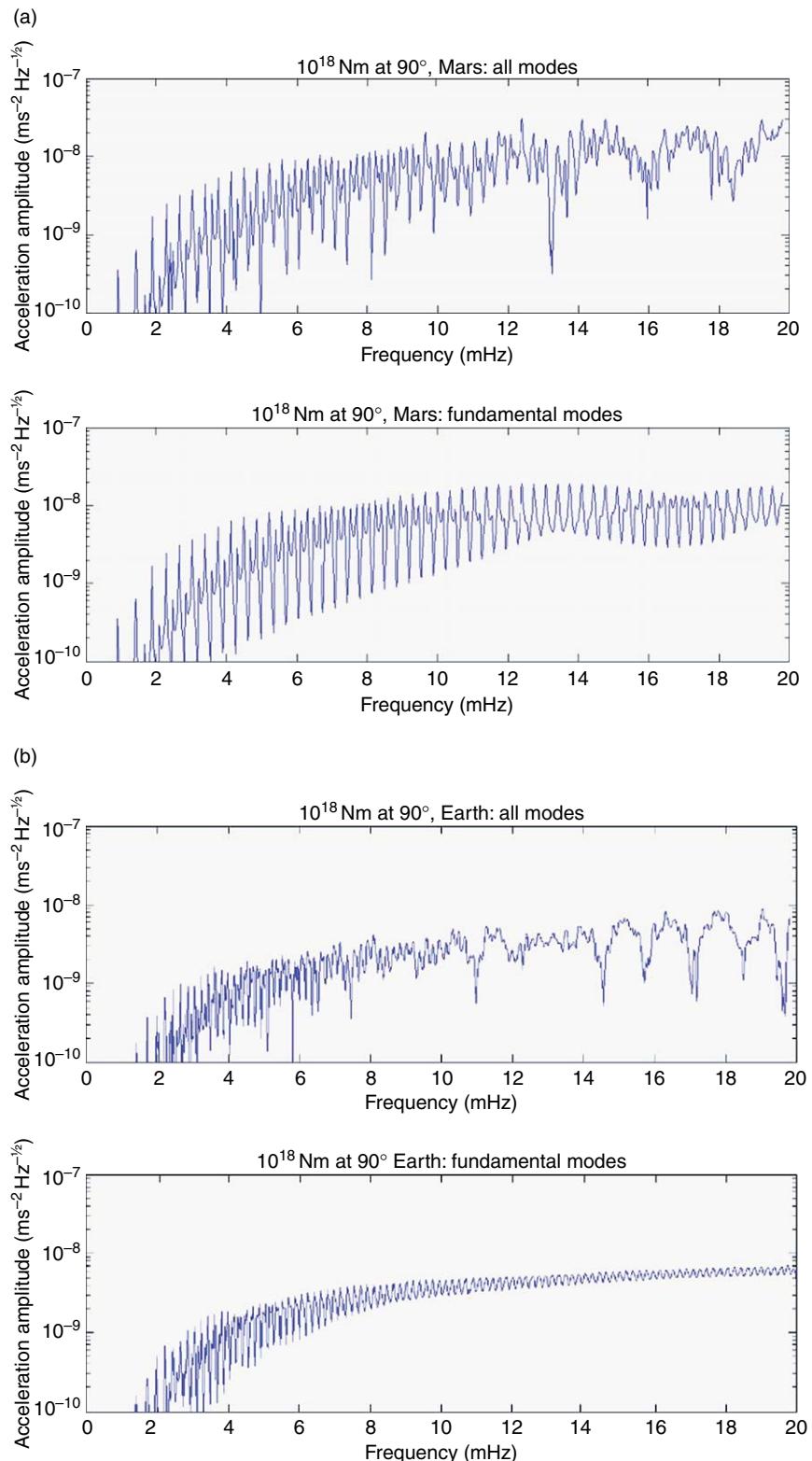


Figure 35 Acceleration amplitude spectrum for a quake with strike, deep, rake angles of 45° , 45° , 45° , and a seismic moment of 10^{18} Nm. Azimuth is 20° and the epicentral distance is 90° . The depth of the event is 30 km. Spectra for Mars are shown in (a). Those for the Earth are in (b). At 5 mHz, the amplitudes exceed $1 \text{ ng Hz}^{-1/2}$ and are a factor 4 greater than those observed on the Earth. Note that below 10 mHz, the fundamental modes account for most of the signal.

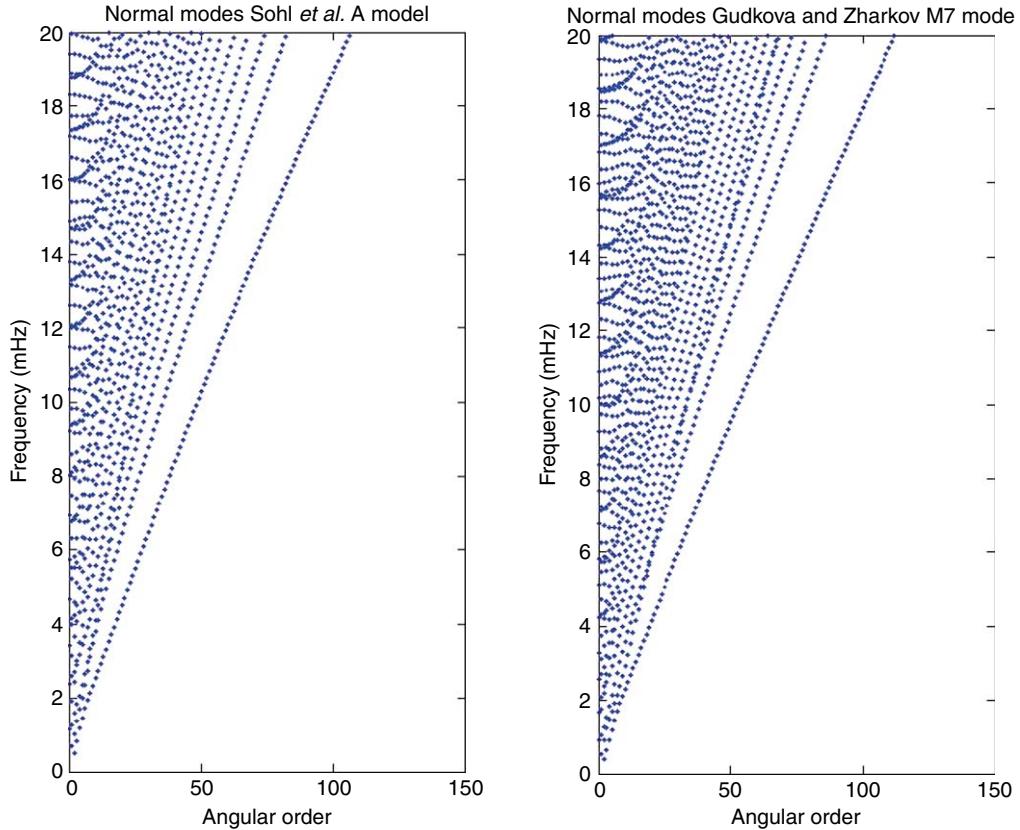


Figure 36 Normal modes for two Mars model (model A from Sohl and Spohn and model M7 from Gudkova and Zharkov (2004)). Differences in the frequencies are about 5–10%.

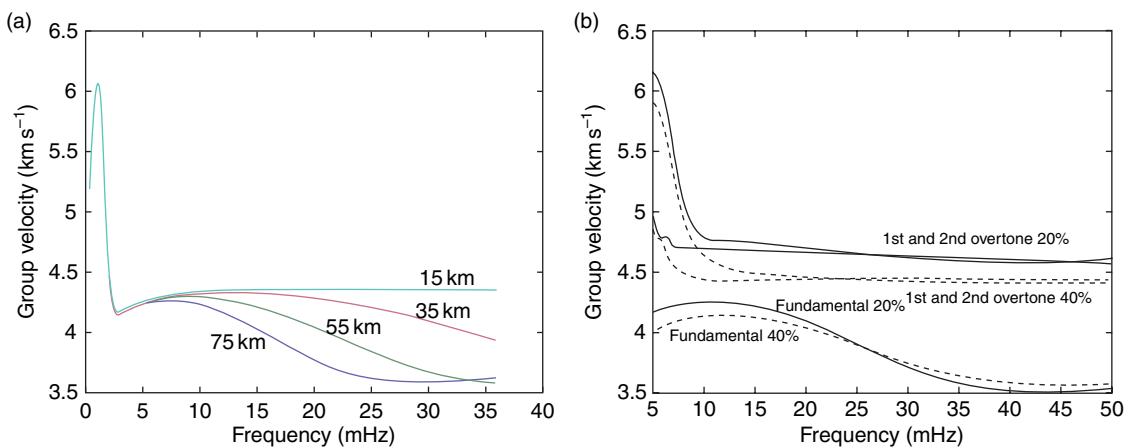


Figure 37 (a) Sensitivity of the Rayleigh waves to the crustal thickness. Model B of Sohl and Spohn (1997) is used and velocity and density in the crust, assumed to be homogeneous, are those of this model. Variations of 10% in the group velocity are found for about 20 km of crustal thickness variation for frequencies in the range of 10–30 mHz. (b) Group velocities for two models of Mocquet et al. (1996) and for the fundamental, first, and second overtones. The iron content of the mantle is indicated in percent. Note that we have added to these models a crust of 50 km thickness, using the Sohl and Spohn (1997) crustal structure. Group velocities without this crust can be found in Lognonné (2005).

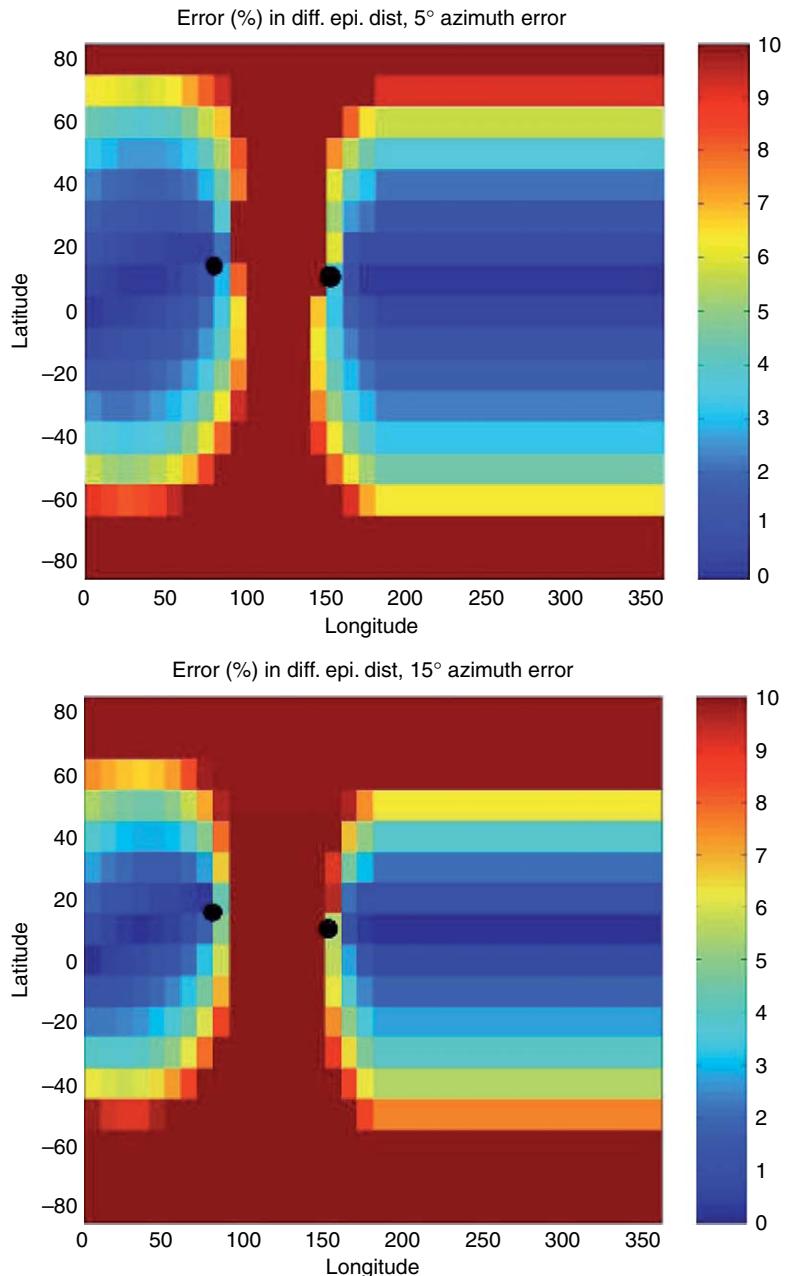


Figure 38 Relative error in the differential epicentral distance for surface-waves analysis based on the back-azimuth determination in two cases. The differential epicentral distance is the difference of epicentral distances between the epicenter and the two stations. Left is for a 5° error, right is for a 15° error. The location of the two stations is 11° N, 152° W and 15° N, 80° W, corresponding to two landing sites proposed for the NetLander Network mission. Only locations with an error smaller than 10% are shown, the latter limit being the *a priori* difference between the seismic values proposed by the models shown on Figure 32.

success of the Apollo seismic network has shown that a planet much less active than the Earth might nevertheless, due to the very low noise level, allow the detection of several events per day. Both meteorite impacts and moonquakes were recorded

on the Moon over a period of more than 7 years. Deep moonquakes remain enigmatic: repeated brittle failure at specific source locations is inferred on the basis of waveform similarity and tidal periodicities. Major questions remain however – tidally

generated stresses are small relative to those expected to result in failure, and why brittle failure occurs at depths of 800–1000 km is not well understood. Large uncertainties in many source region locations exist; to date it has not been possible to determine focal mechanisms for these events, although the prominence of shear-wave arrivals in the seismograms suggests that shear failure is occurring.

Meteorites and moonquakes, together with known artificial impacts have allowed first-order investigations of the variation of seismic velocity with depth. Crustal thickness estimates published shortly after the Apollo missions were suggested to be on the order of 60 km, more recent studies have revised these estimates to 30–45 km. Seismic velocity estimates to depths of ~1000 km are possible; source-receiver geometries prohibit direct seismic investigation of lowermost mantle and core structure. Differences among current model estimates for mantle seismic velocities reflect a combination of network geometry, instrument limitations, and likely 3-D structure. Thus the seismic exploration of the Moon is far from complete. Large uncertainties remain in the internal structure and major discontinuities (core/mantle, lower/upper mantle, mantle crust) are still not well constrained. Moreover, currently unmapped geographical variations in crustal and mantle structure are a critical key to understanding early lunar evolution.

The work reviewed in this chapter demonstrates that the technology exists to perform ground-based seismic experiments on the Moon and on Mars. The installation of a seismic network on Mars has been recommended repeatedly (NRC, 2006); understanding the global internal structure is crucial complementary information to the extensive satellite and rover-based imaging currently underway. Characterization of local, regional, and global seismicity is essential to understanding the present-day tectonic, volcanic, and climatic environment of the planet.

Exciting possibilities also exist for remote (satellite)-based detection of seismicity and seismic waves, via atmospheric sounding. This could prove fruitful in future studies of Venus, for which long-lived ground-based observations are severely challenging due to surface conditions. In addition it could allow investigations of the giant planets.

Planetary seismology is therefore still virgin territory for future planetary investigation. The deployment of planetary seismic networks will be

mandatory for determining the internal structure of telluric planets – critical not only for understanding these bodies, but for comparison with, and increased understanding of, Earth. Satellite-based observations may afford new opportunities in the seismic exploration of solar system bodies with atmospheres, providing otherwise unavailable avenues for investigating Venus and several outer solar system planets and satellites.

Acknowledgments

Catherine Johnson acknowledges support from NASA grants NASA NNG05GK34G, and NAG5-11563, and from IPGP. Philippe Lognonné acknowledges support from CNES, PNP, and INSU/CNRS for a continuous support in the last 10 years. This is IPGP contribution 2221.

References

- Acuña MH, Connerney JEP, Wasilewski P, et al. (1998) Magnetic field and plasma observations at Mars: Initial results of the Mars global surveyor Mission. *Science* 279 (5357): 1676–1680.
- A'Hearn MF, Belton MJS, Delamere WA, et al. (2005) Deep impact: Excavating comet Tempel 1. *Science* 310: 258–264.
- Aki K (1992) Scattering conversions P to S versus S to P. *Bulletin of the Seismological Society of America* 82: 41969–41972.
- Aki K and Richards PG (1980) *Quantitative seismology*. San Francisco, CA: Freeman.
- Anderson DL, Duennebier FK, Latham GV, et al. (1977b) Viking seismology experiment. *Bulletin of the American Astronomical Society* 9: 447.
- Anderson DL and Given J (1982) Absorption band Q model for the Earth. *Journal of Geophysical Research* 87: 3893–3904.
- Anderson DL, Miller WF, Latham GV, et al. (1977a) Seismology on Mars. *Journal of Geophysical Research* 82: 4524–4546.
- Abercrombie Rachel E and Göran E (2001) Earthquake slip on oceanic transform faults. *Nature* 410: 74–77.
- Artru J, Farges T, and Lognonné P (2004) Acoustic waves generated from seismic surface waves: Propagation properties determined from doppler sounding observation and normal-modes modeling 158: 1067–1077.
- Artru J, Lognonné P, and Blanc E (2001) Normal modes modeling of post-seismic ionospheric oscillations. *Geophysical Research Letters* 28: 697–700.
- Assessment of NASA's Mars Architecture, 2007–2016, National Research Council of the National Academies.
- Ball AJ, Lognonné P, Seiferlin K, Spohn T, and Zarnecki JC (2003) Lander and penetrator science for NEO mitigation studies. In: Belton MJS, Yeomans DK, and Morgan TH (eds.) *Mitigation of Hazardous Impacts due to Asteroids and Comets*, ch. 13. Cambridge, UK: Cambridge University Press (ISBN: 0521827647).

- Balmino G, Duron J, Marty J (2005) A new martian gravity field mean model and its time variations. *American Geophysical Union, Fall Meeting 2005*. #G53C.
- Banerdt WB, Chui T, Herrin ET, Rosenbaum D, and Teplitz VL (2006) Lunar seismic search for strange quark matter 37: 1889–1893.
- Beauduin R, Lognonné P, Montagner JP, Cacho S, Karczewski JF, and Morand M (1996) The effect of the atmospheric pressure changes on seismic signals or how to improve the quality of a station. *Bulletin of the Seismological Society of America* 86: 1760–1769.
- Beer T (1974) *Atmospheric Waves*, 315 pp. New York: Halsted Press.
- Benioff H, Press F, and Smith SW (1961) Excitation of the free oscillations of the Earth by earthquake. *Journal of Geophysical Research* 66: 605–620.
- Belleguic V, Lognonné P, and Wieczorek M (2005) Constraints on the martian lithosphere from gravity and topography data. *Journal of Geophysical Research* 110: E11005 (doi:10.1029/2005JE002437).
- Bills BG and Ferrari AJ (1977) A lunar density consistent with topographic, gravitational, librational and seismic data. *Journal of Geophysical Research* 82: 1306–1314.
- Bills BG, Neumann GA, Smith DE, and Zuber M (2005) Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110: E07004 (doi:10.1029/2004JE002376).
- Bills BG and Rubincam DP (1995) Constraints on density models from radial moments: applications to Earth, Moon, and Mars. *Journal of Geophysical Research* 100: 26305–26316.
- Bois E, Boudin F, and Journet A (1996) Secular variation of the Moon's rotation rate. *Astronomy and Astrophysics* 314: 989–994.
- Bratt SR, Bergman EA, and Solomon SC (1985) Thermoelastic stress: How important as a cause of earthquakes in young oceanic lithosphere. *Journal of Geophysical Research* 90: 10249–10260.
- Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics on mars: Evidence from magnetic field history and crust evolution. *Journal of Geophysical Research Planets* 108: 5072 (doi:10.1029/2002JE001999).
- Breuer D and Spohn T (2006) Viscosity of the Martian mantle and its initial temperature: Constraints from crust formation history and the evolution of the magnetic field. *Planetary and Space Science* 54: 153–169.
- Bullen KE (1947) *An Introduction to the Theory of Seismology*. Cambridge, UK: Cambridge University Press.
- Bulow R, Johnson CL, and Bills BG (2006) Tidal stress and deep moonquakes. LPSC 37th, Abstract 1183.
- Bulow R, Johnson CL, Bills BG, and Shearer PM (2007) Temporal and spatial properties of some deep moon quake clusters. *Journal of Geophysical Research* (in press).
- Bulow R, Johnson CL, and Shearer P (2005) New events discovered in the lunar Apollo seismic data. *Journal of Geophysical Research Planets* 110: E10003 (doi:10.1029/2005JE002414).
- Cailleau B, Walter TR, Janle P, and Hauber E (2003) Modeling volcanic deformation in a regional stress field: Implications for the formation of graben structures on alba patera, Mars. *Journal of Geophysical Research Planets* 108(E12): 5141 (doi:10.1029/2003JE002135).
- Cameron AG (2000) Higher-resolution simulations of the giant impact. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 133–144. Tucson, AZ: University of Arizona Press.
- Canup RM and Asphaug E (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.
- Calais E and Minster JB (1995) GPS detection of ionospheric perturbations following the January 17, 1994, northridge earthquake. *Geophysical Research Letters* 22: 1045–1048.
- Cacciani A, Dolci M, Moretti PF, et al. (2001) Search for global oscillations on Jupiter with a double-cell sodium magneto-optical filter. *Astronomy and Astrophysics* 372: 317–325.
- Chenet H, Lognonné P, Wieczorek M, and Mizutani H (2006) Lateral variations of lunar crustal thickness from Apollo seismic dataset. *Earth and Planetary Science Letters* 243: 1–14.
- Chyba CF, Thomas PJ, and Zahnle KJ (1993) The 1908 Tunguska explosion – Atmospheric disruption of a stony asteroid. *Nature* 361: 40–44.
- Cisowski SM, Collinson DW, Runcorn SK, and Stephenson A (1983) A review of lunar paleointensity data and implications for the origin of lunar magnetism. *Journal of Geophysical Research* 88: A691–A704.
- Davis K and Baker DM (1965) Ionospheric effects observed around the time of the alaskan earthquake of March 28, 1964. *Journal of Geophysical Research* 70: 1251–1253.
- Davis PM (1993) Meteoroid impacts as seismic sources on Mars. *Icarus* 105: 469–478.
- Dainty AM and Toksöz NM (1977) Elastic wave propagation in a highly scattering medium: A diffusion approach. *Journal of Geophysics* 43: 375–388.
- Deming D, Mumma MJ, Espenak F, et al. (1989) A search for p-mode oscillations of jupiter~: Serendipitous observations of non-acoustic thermal wave structure. *Astrophysical Journal* 343: 456–467.
- De Pater I and Lissauer JJ (2001) *Planetary Sciences*, 66 pp. Cambridge, UK: Cambridge University Press.
- Dragoni M and Piombo A (2003) A model for the formation of wrinkle ridges in volcanic plains on Venus. *Physics of Earth and Planetary Interiors* 135: 161–171.
- Dreibus G and Wanke H (1985) A volatile-rich planet. *Meteoritics* 20: 367–382.
- Drossart P, Piccioni G, Coradini A, et al. (2004) VIRTIS imaging spectrometer for the ESA/venus express mission, Earth observing systems IX. *Proceedings of the SPIE* 5543: 175–185.
- Ducic V, Artru J, and Lognonné P (2003) Ionospheric remote sensing of the Denali earthquake rayleigh surface waves. *Geophysical Research Letters* 30(18): 1951 (doi:10.1029/2003GL017812).
- Duennebier F and Sutton GH (1974) Thermal moonquakes. *Journal of Geophysical Research* 79: 4351–4364.
- Dziewonski A and Anderson DL (1981) Preliminary reference Earth model. *Physics of Earth and Planetary Interiors* 25: 297–356.
- Dziewonski AM, Hager BH, and O'Connell RJ (1977) Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research* 82: 239–255.
- Ewing M, Latham G, Press F, et al. (1971) Seismology of the Moon and implications on internal structure, origin and evolution. In: Jager D (ed.) *Highlights of Astronomy*, pp. 155–172. Dordrecht: Reidel.
- Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars pathfinder. *Science* 278: 1749–1751.
- Forget F, Hourdin F, Fournier R, et al. (1999) Improved general circulation models of the Martian atmosphere from the surface to above 80 km. *Journal of Geophysical Research* 104: 24155–24176.
- Fuller M and Stanley MC (1987) Lunar paleomagnetism. *Geomagnetism* 2: 307–455.

- Gagnepain-Beyneix J, Lognonné P, Chenet H, and Spohn T (2006) Seismic models of the Moon and their constraints on the mantle temperature and mineralogy. *Physics of Earth and Planetary Interiors* 159: 140–166.
- Garcia R, Crespon F, Ducic V, and Lognonné P (1995) 3D ionospheric tomography of post-seismic perturbations produced by the Denali earthquake from GPS data. *Geophysical Journal International* 163: 1049–1064.
- Garcia R, Lognonné P, and Bonnin X (2005) Detecting atmospheric perturbations produced by Venus quakes. *Geophysical Research Letters* 32: L16205.
- Goins NR, Dainty AM, and Toksöz MN (1981a) Seismic energy release of the Moon. *Journal of Geophysical Research* 86: 378–388.
- Goins NR, Dainty AM, and Toksöz MN (1981b) Lunar seismology: The internal structure of the Moon. *Journal of Geophysical Research* 86: 5061–5074.
- Goins NR and Lazarewicz AR (1979) Martian seismicity. *Geophysical Research Letters* 6: 368–370.
- Golombek MP, Banerdt WB, Tanaka KL, and Tralli DM (1992) A prediction of Mars seismicity from surface faulting. *Science* 258: 979–981.
- Gough DO (1986) NATO ASI Series. vol. C169: *Seismology of the Sun and the Distant Stars; Proceedings of the NATO Advanced Research Workshop*, 491 pp. Dordrecht: D. Reidel Publishing.
- Grimm RE and Hess PC (1997) The crust of Venus. In: Bouger SW, Hunter DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar wind environment*, 1205pp. Tucson, AZ: University of Arizona Press.
- Gudkova T, Mosser B, Provost J, Chabrier G, Gautier D, and Guillot T (1995) Seismological comparison of giant planet interior models. *Astronomy and Astrophysics* 303: 594–603.
- Gudkova TV and Zharkov VN (1997) Models of Jupiter and Saturn with water-depleted atmospheres. *Astronomicheskii Vestnik* 31: 99.
- Gudkova TV and Zharkov VN (1998) The free oscillations of Jupiter. *Planetary and Space Science* 47: 1211–1224.
- Gudkova TV and Zharkov VN (2004) Mars: Interior structure and excitation of free oscillations. *Physics of Earth and Planetary Interiors* 142: 1–22.
- Guillot T (2005) The interiors of giant planets: Models and outstanding questions. *Annual Review of Earth and Planetary Sciences* 33: 493–530.
- Gusev GA and Kurlachev MI (1976) The possibility of observing the lunar microseismic background. *Akademii Nauk SSSR, Izvestiya, Fizika Zemli*. (In Russian.) (Dec) 84–86.
- Hammel HB, Beebe RF, Ingersoll AP, et al. (1995) HST imaging of atmospheric phenomena created by the impact of comet Shoemaker-Levy 9. *Science* 267: 1288–1296.
- Hanks TC and Kanamori H (1979) A moment magnitude scale. *Journal of Geophysical Research* 84: 2348–2350.
- Harder H and Christensen UR (1996) A one plume model of Martian mantle convection. *Nature* 380: 507–509.
- Hood LL (1995) Frozen fields. *Earth, Moon and Planets* 67: 131–142.
- Hood LL (1986) Geophysical constraints on the lunar interior. In: Hartmann WK, Philips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 361–410. Huston, TX: Lunar and Planetary Institute.
- Hood LL and Jones JH (1987) Geophysical constraints on lunar bulk composition and structure: A reassessment. *Journal of Geophysical Research* 92: E396–E410.
- Hood LL, Mitchell DL, Lin RP, Acuna MH, and Binder AB (1999) Initial measurements of the lunar induced magnetic dipole moment using lunar prospector magnetometer data. *Geophysical Research Letters* 26: 2327–2330.
- Hood LL and Zuber MT (2000) 1986 Recent refinements in geophysical constraints on Lunar origin and evolution. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 397–409. Houston, TX: Lunar and Planetary Institute.
- Hunten DM, Colin L, and Donahue TM (1983) *Venus*, 1143 pp. Tucson, AZ: University of Arizona Press.
- Ingersoll AP and Kanamori H (1995) Waves from the collisions of comet Shoemaker-Levy-9 with Jupiter. *Nature* 374: 706–708.
- Ingersoll AP, Kanamori H, and Dowling TE (1994) Atmospheric gravity waves from the impact of comet Shoemaker-Levy 9 with Jupiter. *Geophysical Research Letters* 21: 1083–1086.
- Jolliff BL, Gillis JJ, Haskin LA, Korotev RL, and Wieczorek MA (2000) Major lunar crustal terranes: Surface expression and crust-mantle origins. *Journal of Geophysical Research* 105: 4197–4216.
- Johnson CL, Stixrude L, Lithgow-Bertelloni C, Bulow RC, and Shearer PM (2005) Mineralogical and seismological models of the lunar mantle. 36th Annual Lunar and Planetary Science Conference, March 14–18, 2005, League City, TX, 1565.
- Kanamori H (1993) Excitation of jovian normal modes by an impact source. *Geophysical Research Letters* 20: 2921–2924.
- Kanamori H (2004) Some fluid-mechanical problems in geophysics-wave in the atmosphere and fault lubrication. *Fluid Dynamics Research* 34: 1–19.
- Kanamori H and Mori J (1992) Harmonic excitation of mantle Rayleigh waves by the 1991 eruption of Mount Pinatubo, Philippines. *Geophysical Research Letters* 19: 721–724.
- Kanamori H, Mori J, and Harkrider DG (1994) Excitation of atmospheric oscillations by volcanic eruptions 22: 21947–21961.
- Kaula WM (1979) The moment of inertia of Mars. *Geophysical Research Letters* 6: 194–196.
- Kelley M, Livingston R, and McCready M (1985) Large amplitude thermospheric oscillations induced by earthquakes. *Geophysical Research Letters* 12: 577–580.
- Khan A, MacLennan J, Taylor SR, and Connolly JAD (2006) Are the Earth and the Moon compositionally alike? inferences on lunar composition and implications for lunar origin and evolution from geophysical modeling. *Journal of Geophysical Research* 111: E05005 (doi:10.1029/2005JE002608).
- Khan A and Mosegaard K (2001) New information on the deep lunar interior from an inversion of lunar free oscillation periods. *Geophysical Research Letters* 28: 1791–1794.
- Khan A and Mosegaard K (2002) An inquiry into the lunar interior: a non linear inversion of the Apollo seismic data. *Journal of Geophysical Research* 107, doi:10.1029/2001JE001658.
- Khan A, Mosegaard K, and Rasmussen KL (2000) A new seismic velocity model for the Moon from a Monte Carlo inversion of the Apollo lunar seismic data. *Geophysical Research Letters* 27: 1591–1594.
- Khan A, Mosegaard K, Williams JG, and Lognonné P (2004) Does the Moon possess a molten core? probing the deep lunar interior using results from LLR and lunar prospector. *Journal of Geophysical Research* 109: E09007 (doi:10.1029/2004JE002294).
- Knapmeyer M, Oberst J, Hauber E, Wählisch M, Deuchler C, and Wagner R (2006) Implications of the Martian surface fault distribution and lithospheric cooling for seismicity: A working model. *Journal of Geophysical Research* 111: E11006 (doi:10.1029/2006JE002708).
- Kobayashi N and Nishida K (1998) Continuous excitation of planetary free oscillations by atmospheric disturbances. *Nature* 395: 357–360.

- Konopliv AS, Binder AB, Hood LL, Kucinskas AB, Sjogren WL, and Williams JG (1998) Improved gravity field of the Moon from lunar prospector. *Science* 281: 1476–1479.
- Koyama J and Nakamura Y (1980) Focal mechanism of deep moonquakes. *Proceedings of the Lunar and Planetary Science Conference* 11th: 1855–1855.
- Kovach RL and Chyba CF (2001) Seismic detectability of a subsurface ocean on Europa. *Icarus* 150: 279–287.
- Kovach RL and Watkins JS (1973a) Apollo 17 seismic profiling—probing the lunar crust. *Science* 180: 1063–1064.
- Kovach RL and Watkins JS (1973b) The structure of the lunar crust at the Apollo 17 site. *Geochimica et Cosmochimica Acta* 37 (Suppl. 4): 2549–2560.
- Khavroskin OB and Tsyplakov VV (1996) Penetrator «Mars-96», reality and possibilities of seismic experiment, in experimental problems in planetology. 1 *United Institute of Physics of the Earth*, Russian Academy of Science, Moscow.
- Kochan H, Feibig W, Konopka U, et al. (2000) CASSE – The ROSETTA lander comet acoustic surface sounding experiment – Status of some aspects, the technical realisation and laboratory simulations. *Planetary and Space Science* 48: 385–399.
- Konopliv AS, Binder AB, Hood LL, Kucinskas AB, Sjogren WL, and Williams JG (1998) Improved gravity field of the Moon from lunar prospector. *Science* 281: 1476–1480.
- Kuramoto K (1997) Accretion, core formation, H and C evolution of the Earth and Mars. *Physics of Earth and Planetary Interiors* 100: 3–20.
- Kuskov OL (1995) Constitution of the Moon. Part 3: Composition of middle mantle from seismic data. *Physics of Earth and Planetary Interiors* 102: 239–257.
- Kuskov OL (1997) Constitution of the Moon. Part 4: Composition of the mantle from seismic data. *Physics of Earth and Planetary Interiors* 90: 55–74.
- Kuskov OL and Fabrichanaya OB (1994) Constitution of the Moon. Part 2: Composition and seismic properties of the lower mantle. *Physics of Earth and Planetary Interiors* 83: 197–216.
- Kuskov OL and Kronrod VA (1998) Constitution of the Moon. Part 5: Constraints on composition, density, temperature and radius of a core. *Physics of Earth and Planetary Interiors* 107: 285–306.
- Kuskov OL and Kronrod VA (2000) Resemblance and difference between the constitution of the Moon and Io. *Planetary and Space Science* 48: 717–726.
- Kuskov OL, Kronrod VA, and Hood LL (2002) Geochemical constraints on the seismic properties of the lunar mantle. *Physics of Earth and Planetary Interiors* 134: 175–189.
- Kuskov OL, Kronrod VA, and Hood LL (2003) Geochemical constraints on the seismic properties of the lunar mantle. *Physics of Earth and Planetary Interiors* 134: 175–189.
- Lammlein DR (1973) *Lunar Seismicity, Structure and Tectonics*. PhD. Dissertation, Columbia University, New York, 99 pp.
- Lammlein D (1977) Lunar seismicity and tectonics. *Physics of Earth and Planetary Interiors* 14: 224–273.
- Lammlein D, Latham GV, Dorman J, Nakamura Y, and Ewing M (1974) Lunar seismicity, structure and tectonics. *Reviews of Geophysics and Space Physics* 12: 1–21.
- Larmat C, Montagner JP, Capdeville Y, Banerdt WB, Lognonné P, and Viloteau JP (2006) Numerical assessment of the effects of topography and crustal thickness on Martian seismograms using a coupled modal solution-spectral elements method. (*Icarus*) (submitted).
- Larose E, Khan A, Nakamura Y, and Campillo M (2005) Lunar subsurface investigated from correlation of seismic noise. *Geophysical Research Letters* 32: L16201.
- Latham GV, Ewing M, Dorman J, et al. (1970b) Seismic data from man-made impacts on the Moon. *Science* 170: 620–626.
- Latham GV, Ewing M, Dorman J, et al. (1971) Moonquakes. *Science* 174: 687–692.
- Latham GV, Ewing M, Press F, et al. (1970a) Passive seismic experiment. *Science* 167: 455–457.
- Latham G, Ewing M, Press F, and Sutton G (1969) The apollo passive seismic experiment. *Science* 165: 241–250.
- Lederer SM, Marley MS, Mosser B, Maillard JP, Chanover NJ, and Beebe RF (1995) Albedo features and jovian seismology. *Icarus* 114: 269–277.
- Lee S, Zanolin M, Thode AM, Pappalardo RT, and Makris NC (2003) Probing Europa's interior with natural sound sources. *Icarus* 165: 144–167.
- Lehner FE, Witt EO, Miller WF, and Gurney RD (1962) A seismometer for lunar experiments. *Journal of Geophysical Research* 67: 4779–4786.
- Leonard RS and Barnes RA, Jr. (1965) Observation of ionospheric disturbances following the alaska earthquake. *Journal of Geophysical Research* 70: 1250.
- Libbrecht KG and Woodard MF (1991) Advances in helioseismology. *Science* 253: 152–157.
- Linkin V, Harri AM, Lipatov A, et al. (1998) A sophisticated lander for scientific exploration of Mars: Scientific objectives and implementation of the Mars96 small station. *Planetary Space Sciences* 46: 717–737.
- Lodders K (2000) An oxygen isotope mixing model for the accretion and composition of rocky planets. *Space Science Reviews* 92: 341–354.
- Lodders K and Fegley B (1997) An oxygen isotope model for the composition of Mars. *Icarus* 126: 373–379.
- Lognonné P (2005) Planetary seismology. *Annual Review of Earth and Planetary Sciences* 33: 19.1–19.34 (doi:10.1146/annurev.earth.33.092203.122605).
- Lognonné P, Artru J, and Forget F (2007) Simulation of the continuous excitation of normal modes of Mars (abstract). Nantes: 2nd Netlander Scientific Workshop.
- Lognonné P, Artru J, Garcia R, et al. (2006) Ground based GPS tomography of ionospheric post-seismic signal during demeter: The SPECTRE project. *Planetary and Space Science* 54: 528–540.
- Lognonné P and Clévidé E (2002) Normal modes of the Earth and Planets. In: Kanamori H, Jennings P, and Lee W (eds.) *International Geophysical Series 81A: Handbook on Earthquakes and Engineering Seismology*, pp. 125–147. San Diego: IASPEI Centennial Publications, Academic Press.
- Lognonné P, Clévidé E, and Kanamori H (1998b) Normal mode summation of seismograms and barograms in an spherical Earth with realistic atmosphere. *Geophysical Journal International* 135: 388–406.
- Lognonné P, Gagnepain-Beyneix J, Banerdt WB, Cacho S, Karczewski JF, and Morand M (1996) An ultra-broad band seismometer on inter marsnet. *Planetary Space Sciences* 44: 1237–1249.
- Lognonné P, Giardini D, Banerdt B, et al. (2000) The NetLander very broad band seismometer. *Planetary and Space Science* 48: 1289–1302.
- Lognonné P and Mosser B (1993) Planetary seismology. *Survey in Geophysics* 14: 239–302.
- Lognonné P, Mosser B, and Dahmen FA (1994) Excitation of the jovian seismic waves by the Shoemaker-Levy 9 cometary impact. *Icarus* 110: 186–195.
- Lognonné P, Zharkov VN, Karczewski JF, et al. (1998a) The seismic optimism experiment. *Planetary Space Sciences* 46: 739–747.
- Lognonné P, Gagnepain-Beyneix J, and Chenet H (2003) A new seismic model of the Moon: Implication in terms of structure, formation and evolution. *Earth and Planetary Science Letters* 6637: 1–18.

- Lognonné P (2005) Planetary seismology. *Annual Review of Earth and Planetary Sciences* 33: 571–604 (doi:10.1146/annurev.earth.33.092203.122605).
- Loudin MG and Alexander SS (1978) Observed estimates of spheroidal free oscillations of the Moon and their interpretation. *Eos Transaction, American Geophysical Union* 59: 1124.
- Mackwell SJ, Zimmerman ME, and Kohlstedt DL (1998) High-temperature deformation of dry diabase with application to tectonics on venus. *Journal of Geophysical Research* 103: 975–984.
- Marley MS (1994) Seismological consequences of the impact of comet Shoemaker-Levy 9. *Astrophysical Journal* 427: L63–L66.
- Marley S and Porco CC (1993) Planetary acoustic mode seismology – Saturn's rings. *Icarus* 106: 508–524.
- Miller WF (1963) The caltech digital seismograph. *Journal of Geophysical Research* 68(3): 841–847.
- Minshull TA and Goult NR (1988) The influence of tidal stresses on deep moonquake activity. *Physics of Earth and Planetary Interiors* 53: 41–55.
- Mocquet A (1998) A search for the minimum number of stations needed for seismic networking on Mars. *Planetary and Space Science* 47: 397–409.
- Mocquet A, Vacher P, Grasset O, and Sotin C (1996) Theoretical seismic models of Mars: The importance of the iron content of the mantle. *Planetary and Space Science* 44: 1251–1268.
- Mosser B (1995) Propagation and trapping of global oscillations in the Jovian troposphere and stratosphere. *Astronomy and Astrophysics* 293: 586–593.
- Mosser B (1996) Giant planets seismology in sounding solar and stellar interiors. In: Provost J and Schmider F-X (eds.) *Proceedings of the 181st Symposium of the International Astronomical Union*, pp. 251–264. Dordrecht: Kluwer Academic Publishers.,
- Mosser B, Galdemard P, Lagage P, et al. (1996) Impact seismology: A search primary pressure waves following impacts A and H. *Icarus* 121: 331–340.
- Mosser B, Maillard JP, and Mekarnia D (2000) New attempt at detecting the jovian oscillations. *Icarus* 144: 104–113.
- Mosser B, Mekarnia D, Maillard JP, Gay J, Gautierand D, and Delache P (1993) Seismological observations with a Fourier transform spectrometer – Detection of jovian oscillations. *Astronomy and Astrophysics* 267: 604–622.
- Mosser B, Schmider F-X, Delache P, and Gautier D (1991) A tentative identification of jovian global oscillations. *Astronomy and Astrophysics* 251: 356–364.
- Mueller S, Taylor GJ, and Phillips RJ (1988) Lunar composition – A geophysical and petrological synthesis. *Journal of Geophysical Research* 93: 6338–6352.
- Munk WH and MacDonald GJF (1960) *The Rotation of the Earth: A Geophysical Discussion*. New York: Cambridge University Press.
- Nakamura Y and Anderson DL (1979) Martian wind activity detected by a seismometer at Viking Lander 2 site. *Geophysical Research Letters* 6: 499–502.
- Nakamura Y (1976) Seismic energy transmission in the lunar surface zone determined from signals generated by movement of lunar rovers. *Bulletin of the Seismological Society of America* 66: 593–606.
- Nakamura Y (1977) Seismic energy transmission in an intensely scattering environment. *Journal of Geophysical Research* 43: 389–399.
- Nakamura Y (1978) A1 Moonquakes: Source distribution and mechanisms. *Proceedings of the Lunar Science Conference* 9th: 3589–3607.
- Nakamura Y (1983) Seismic velocity structure of the lunar mantle. *Journal of Geophysical Research* 88: 677–686.
- Nakamura Y (2003) New identification of deep moonquakes in the Apollo lunar seismic data. *Physics of Earth and Planetary Interiors* 139: 197–205.
- Nakamura Y (2005) Farside deep moonquakes and deep interior of the Moon. *Journal of Geophysical Research* 110: E01001 (doi:10.1029/2004JE002332).
- Nakamura Y, Dorman J, Duennebier F, Lammlein D, and Latham G (1975) Shallow lunar structure determined from the passive seismic experiment. *Moon* 13: 57–66.
- Nakamura Y, Duennebier FK, Latham GV, and Dorman HJ (1976) Structure of the lunar mantle. *Journal of Geophysical Research* 81: 4818–4824.
- Nakamura Y, Latham GV, and Dorman HJ (1980) How we processed apollo lunar seismic data. *Physics of Earth and Planetary Interiors* 21: 218–224.
- Nakamura Y, Latham GV, and Dorman HJ (1982) Apollo lunar seismic experiment- final summary. *Proceedings of the Lunar and Planetary Science Conference* 87: A117–A123.
- Nakamura Y and Koyama J (1982) Seismic Q of the lunar upper mantle. *Journal of Geophysical Research* 87: 4855–4861.
- Nakamura Y, Latham GV, Dorman J, and Harris J (1981) Passive seismic experiment long-period event catalog, final version, *Technical Report* No 18, Galveston Geophysics Laboratory, University of Texas at Austin.
- Nakamura Y, Latham GV, Dorman HJ, and Harris JE (2004) Passive seismic experiment long-period event catalog, *Technical Report* 118, Review Edition, University of Texas, Institute for Geophysics, Austin.
- Nakamura Y, Latham G, Lammlein D, Ewing M, Duennebier F, and Dorman J (1974) Deep lunar interior inferred from recent seismic data 1: 137–140.
- Nataf H and Ricard Y (1996) 3SMAC: An *a priori* tomographic model of the upper mantle based on geophysical modeling. *Physics of Earth and Planetary Interiors* 95: 101–122.
- Neukum G, Jaumann R, Hoffmann H, et al. (2004) Recent and episodic volcanic and glacial activity on mars revealed by the high resolution stereo camera. *Nature* 432: 971–979.
- Nimmo F and Schenk P (2005) Normal faulting on europa: Implications for ice shell properties. *Journal of Structural Geology* 28: 2194–2203 (doi:10.1016/j.jsg.2005.08.009).
- Nishida K, Kobayashi N, and Fukao Y (2000) Resonant oscillations between the solid Earth and the atmosphere. *Science* 287: 2244–2246.
- Oldham RD (1906) The constitution of the Earth, quart. *Journal of the Geological Society London* 62: 456–475.
- Oberst J (1987) Unusually high stress drop associated with shallow moonquakes. *Journal of Geophysical Research* 92: 1397–1405.
- Okan E and Anderson DJ (1978) Theoretical models for Mars and their seismic properties. *Icarus* 33: 514–528.
- Panning M, Lekic V, Manga M, Cammarano F, and Romanowicz B (2006) Long-period seismology on Europa 2: predicted seismic response. *Journal of Geophysical Research* 111(E12): E12008.
- Panning M, Lekic V, Manga M, Cammarano F, and Romanowicz B (2007) Long-period seismology on europa: II. predicted seismic response. *Journal of Geophysical Research* 111: E12008 (doi:10.1029/2006JE002712).
- Papanicolaou G, Ryzhik L, and Keller JB (1996) Stability of the P to S-energy ratio in the diffusive regime. *Bulletin of the Seismological Society of America* 86: 1107–1115.
- Peterson J (1993) Observation and modeling of seismic background noise. *US Geological Survey Open-File Report* 93–322: 1–45.
- Phillips RJ (1991) Expected rate of marsquakes in scientific rationale and requirements for a global seismic network on Mars. *LPI Technical Report* 91–02, pp. 35–38. Houston, TX: Lunar and Planetary Institute, Houston.

- Poveda A, Herrera MA, García JL, Hernández-Alcántara A, and Curioca K (1999) The expected frequency of collisions of small meteorites with cars and aircraft. *Planetary and Space Science* 47: 715–719.
- Press F, Buwalda P, and Neugebaue MR (1960) A lunar seismic experiment. *Journal of Geophysical Research* 65: 3097–3105.
- Richardson JE, Melosh HJ, and Greenberg R (2004) Impact-induced seismic activity on asteroid 433 eros: A surface modification. *Science* 306: 1526–1529.
- Righter K (2002) Does the Moon have a metallic core? constraints from giant impact modeling and siderophile elements. *Icarus* 158: 1–13.
- Rhie A and Romanowicz B (2004) Excitations of the Earth's incessant free oscillation by atmosphere/ocean/solid Earth coupling. *Nature* 431: 552–556.
- Rhie J and Romanowicz B (2006) A study of the relation between ocean storms and the Earth's hum. *Geochemistry Geophysics Geosystems* 7: 10 (CitelID Q10004).
- Riedel KS and Sidorenko A (1995) Minimum bias multiple taper spectral estimation. *IEEE Transactions on Signal Processing* 43: 188–195.
- Rubie DC, Gessmann CK, and Frost DJ (2004) Partitioning of oxygen during core formation on the Earth and Mars. *Nature* 429: 58–61.
- Sanloup C, Jambon A, and Gillet P (1999) A simple chondritic model of Mars. *Physics of Earth and Planetary Interiors* 112: 43–54.
- Sanloup C, Guyot F, Gillet P, and Fei Y (2002) Physical properties of liquid Fe alloys at high pressure and their bearings on the nature of metallic planetary cores. *Journal of Geophysical Research* 107, doi:10.1029/2001JB000808.
- Saumon D and Chabrier G (1989) Fluid hydrogen at high density~: The plasma phase transition. *Physical Review Letters* 62: 2397–2400.
- Saumon D, Hubbard WB, Chabrier G, and van Horn HM (1992) The role of the molecular-metallic transition of hydrogen in the evolution of Jupiter, Saturn, and brown dwarfs. *Astrophysical Journal* 391: 827–831.
- Schmider F-X, Mosser B, and Fossat E (1991) A possible detection of jovian global oscillations. *Astronomy and Astrophysics* 248: 281–291.
- Scholz CH (1990) The Mechanism of Earthquakes and Faulting. Cambridge, UK: Cambridge University Press.
- Schmider F-X, Gay J, Jacob C, et al. (2003) SYMPA: A dedicated instrument and a network for seismology of giant planets. *Astrophysics and Space Science* 284: 449–452.
- Shapiro NM and Campillo M (2004) Emergence of broadband rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters* 31: L07614 (doi:10.1029/2004GL019491).
- Shapiro NM, Campillo M, Michel, Rittwoller MH, and Michael H (2005) High-resolution surface-wave tomography from ambient seismic noise. *Science* 307: 1615–1618.
- Shearer P (1999) *Introduction to Seismology*, 27 pp. Cambridge, UK: Cambridge University Press.
- Shoemaker EM, Wolfe RF, and Shoemaker CS (1990) Asteroid and comet flux in the neighborhood of Earth. *Geological Society of America, Special Paper* 247
- Smith DE, Zuber MT, Torrence MH, and Dunn PJ (2003) Estimating the k2 tidal gravity love number of Mars. *American Geophysical Union, Fall Meeting 2003*, P31A-05
- Sohl F and Spohn T (1997) The interior structure of Mars: Implications from SNC Meteoroids. *Journal of Geophysical Research* 102: 1613–1636.
- Sohl F, Schubert G, and Spohn T (2005) Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110, doi:10.1029/2005JE002520.
- Solomon S, Bullock M, and Grinspoon D (1999) Climate change as a regulator of tectonics on Venus. *Science* 286: 87–90.
- Spohn T, Acuña MH, Breuer D, Golombek M, et al. (2001b) Geophysical constraints on the evolution of Mars. *Space Science Reviews* 96: 231–262.
- Spohn T, Konrad W, Breuer D, and Ziethe R (2001a) The longevity of lunar volcanism: Implications of thermal evolution calculations with 2D and 3D mantle convection models. *Icarus* 149: 54–65.
- Steeple DW, Schmeissner C, and Macy B (1997) Recording wind microstructure with a seismograph. *Geophysical Research Letters* 24: 2375–2377.
- Stevenson DJ (2001) Mars' core and magnetism. *Nature* 412: 214–219.
- Stofan ER, Saunders RS, Senske D, Nock K, and Tralli D (1993) Venus interior structure mission (VISM): Establishing a seismic network on Venus. *Lunar and Planetary Institute, Workshop on Advanced Technologies for Planetary Instruments*, Part 1, pp. 23–24, (See N93-28764 11–91). Houston, Texas: Lunar and Planetary Institute.
- Suda N, Nawa K, and Y Fukao Y (1998) Incessant excitation of the Earth's free oscillations. *Science* 279: 2089–2091.
- Surkov Y (1990) *Exploration of Terrestrial Planets from Spacecraft*, 390 p, page 372–373. New York: Ellis Horwood.
- Sutton GH and Latham GV (1964) Analysis of a feedback controlled seismometer. *Journal of Geophysical Research* 69: 3865–3882.
- Sutton GH and Steinbacher R (1967) Surveyor seismograph experiment. *Journal of Geophysical Research* 72: 841–844.
- Tanimoto T (1999) Excitation of normal modes by atmospheric turbulence: Source of long period seismic noise. *Geophysical Journal International* 136: 395–402.
- Tanimoto T (2001) Continuous free oscillations: Atmosphere-solid Earth coupling. *Annual Review of Earth and Planetary Sciences* 29: 563–584.
- Tanimoto T (2005) The oceanic excitation hypothesis for the continuous oscillations of the Earth. *Geophysical Journal International* 160: 276–288 (doi:10.1111/j.1365-246X.2004.02484.x).
- Tanimoto T, Um J, Nishida K, and Kabayashi N (1998) Earth's continuous oscillations observed on seismically quiet days. *Geophysical Research Letters* 25: 101553–101556.
- Taylor SR (1982) *Planetary Science: A Lunar perspective*. Houston, TX: Lunar and Planetary Institute.
- Tibi R and Wiens DA (2005) Detailed structure and sharpness of upper mantle discontinuities in the tonga subduction zone from regional broadband arrays. *Journal of Geophysical Research* 110: B6 (doi:10.1029/2004JB003433).
- Tobias RL (1978) *The Lunar Surface Gravimeter and the Search for Gravitational Radiation*, PhD Thesis, Maryland University, College Park.
- Toksöz MN (1974) Geophysical data and the interior of the Moon. *Annual Review of Earth and Planetary Sciences* 2: 151.
- Toksöz MN, Goins NR, and Cheng CH (1977) Moonquakes: Mechanisms and relations to tidal stresses. *Science* 196: 979–981.
- Tolstoy M, Vernon FL, Orcutt JA, and Wyatt FK (2002) The breathing of the seafloor: Tidal correlations of seismicity on axial volcano. *Geology* 30: 503–506.
- US Standard atmosphere (1976) *Committee on the extension to the standard atmosphere*. Washington, DC: US Government printing office.
- Van Hoolst T, Dehant V, Roosbeek F, and Lognonné P (2003) Tidally induced surface displacements, external potential variations, and gravity variations on Mars. *Icarus* 161: 281–296.
- Van Thienen P, Rivoldini A, van Hoolst T, and Lognonné P (2006) A top-dow origin for Martian mantle plumes. *Icarus* 185: 197–201.

- Verhoeven O, Rivoldini A, Vacher P, et al. (2005) Planetary interiors structure inferred from electromagnetic, geodetic and seismic network science. Part I: Forward problem an the case of Mars. *Journal of Geophysical Research* 110, doi:10.1029/2004JE002271.
- Vinnik L, Chenet H, Gagnepain-Beyneix J, and Lognonné P (2001) First seismic receiver functions on the Moon. *Geophysical Research Letters* 28: 3031–3034.
- Von Rebeur-Paschwitz E (1889) The earthquake of Tokyo, April 18, 1889. *Nature* 40: 294–295.
- Vorontsov SV and Zharkov VN (1981) Natural oscillations of the giant planets~: Influence of rotation and ellipticity. *Astronomy Zeitschrift* 58: 1101–1114 (also in Sov. Astron., 25, 627–634, 1982).
- Vorontsov SV, Zharkov VN, and Lubimov VM (1976) The free oscillations of Jupiter and Saturn. *Icarus* 27: 109.
- Waenke H, Baddehausen H, Blum K, et al. (1977) On the chemistry of lunar samples and achondrites – Primary matter in the lunar highlands: A re-evaluation. Lunar Science Conference, 8th: 22191–131977.
- Walker JD and Huebner WF (2004) Seismological investigation of asteroid and comet interiors. In: Belton MJS, Yeomans DK, and Morgan TH (eds.) *Mitigation of Hazardous Impacts due to Asteroids and Comets*, pp. 234–265. Cambridge, UK: Cambridge University Press (ISBN: 0521827647).
- Walter CM, Marley MS, Hunten DM, et al. (1996) A search for seismic waves from the impact of the SL/9 R Fragment. *Icarus* 121: 341–350.
- Walterscheid RL, Brinkman DG, and Schubert G (2000) Wave disturbances from the comet SL-9 impacts into Jupiter's atmosphere. *Icarus* 145: 140–146.
- Ward SN (1998) On the consistency of earthquake moment rates, geological fault data, and space geodetic strain: The United States. *Geophysical Journal International* 134(1): 172–186.
- Watkins JS and Kovach RL (1972) Apollo 14 active seismic experiment. *Science* 175: 1244–1245.
- Weaver PF, Yuen PC, Proffitt GW, and Furumoto AS (1970) Acoustic coupling in the ionosphere from seismic waves of the earthquake at Kurile Islands on August 11, 1969. *Nature* 226: 1239–1241.
- Webb SC (2007) 'The Earth's Hum' is driven by ocean waves over the continental shelves. *Nature* 445: 754–756 (doi:10.1038/nature05536).
- Weber J (1971) Gravitational radiation experiments. In: Charles GK and Peres A (eds.) *Relativity and Gravitation, Based on the Proceedings of an International Seminar on Relativity and Gravitation*, 309 p. New York: Gordon and Breach.
- Wieczorek MA, Bradley M, Jolliff L, et al. (2006) The constitution and structure of the lunar interior. *Reviews in Mineralogy and Geochemistry* 60: 221–364.
- Wieczorek MA and Phillips RJ (1998) Potential anomalies on a sphere: Applications to the thickness of the lunar crust. *Journal of Geophysical Research* 103: 1715–1724.
- Widmer R and Zürn W (1992) Bichromatic excitation of long-period rayleigh and air waves by the Mount Pinatubo and el chichón volcanic eruptions. *Geophysical Research Letters* 19: 765–768.
- Williams JG, Boggs DH, Yoder CF, Ratcliff JT, Todd J, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106: 27933–27968.
- Woodhouse JH and Dziewonski AM (1984) Mapping of the upper-mantle: Three dimensional modeling of Earth structure by inversion of seismic waveforms. *Journal of Geophysical Research* 89: 5953–5986.
- Yoder CF, Konopliv AS, Yuan DN, Standish EM, and Folkner WM (2003) Fluid core size of Mars from detection of the solar tide. *Science* 300: 299–303.
- Yuen PC, Weaver PF, Suzuki RK, and Furumoto AS (1969) Continuous traveling coupling between seismic waves and the ionosphere evident in may 1968 Japan earthquake data. *Journal of Geophysical Research* 74: 2256–2264.
- Zharkov VN and Gudkova TV (1997) On the dissipative factor of Martian interiors. *Planetary and Space Science* 45: 401–407.
- Zharkov VN and Gudkova TV (2000) Interior structure models, Fe/Si ratio and parameters of figure for Mars. *Physics of Earth and Planetary Interiors* 117: 407–420.
- Zharkov VN and Gudkova TV (2005) Construction of Martian interior model. *Solar System Research* 39: 343–373 (doi: 10.1007/s11208-005-0049-7).
- Zuber MT (2001) The crust and mantle of Mars. *Nature* 412: 220–227.
- Zuber MT, Solomon SC, Phillips RC, et al. (2000) Internal structure and early thermal evolution of Mars from Mars global surveyor topography and gravity. *Science* 287: 1788–1793.
- Zürn W and Widmer R (1996) World wide observation of bichromatic long-period rayleigh-waves excited during the June 15, 1991 eruption of Mt. Pinatubo. In: Newhall C and Punongbayan R (eds.) *Fire and Mud, Eruptions of Mount Pinatubo*, pp. 615–624. Philippines: Philippin Institute of Volcanology and Seismology. Quezo City and University of Washington Press.

10.04 The Rotation of the Terrestrial Planets

T. Van Hoolst, Royal Observatory of Belgium, Brussels, Belgium

© 2007 Elsevier B.V. All rights reserved.

10.04.1	Introduction	124
10.04.2	Theoretical Foundations	125
10.04.2.1	Conservation of Angular Momentum	125
10.04.2.1.1	Precession	125
10.04.2.1.2	Nutation, polar motion, and length-of-day variations	126
10.04.2.1.3	Kinematical relation between polar motion and nutation	127
10.04.2.2	Liouville Equations	129
10.04.2.2.1	Excitation	129
10.04.2.2.2	Basic equations	129
10.04.2.2.3	Core–mantle coupling	131
10.04.2.2.4	Mass redistribution	131
10.04.2.2.5	Periodic rotation variations	132
10.04.2.3	Hamiltonian Equations and Long-Term Evolution	133
10.04.2.4	Dissipative Effects as Causes of Secular Rotation Variations	135
10.04.2.4.1	Gravitational tides	135
10.04.2.4.2	Thermal atmospheric tides	136
10.04.2.4.3	Core–mantle friction	137
10.04.2.4.4	Climate friction	138
10.04.3	Mars	138
10.04.3.1	Long-Term Spin Evolution	138
10.04.3.2	Nutation	139
10.04.3.2.1	Rigid nutations	139
10.04.3.2.2	Nonrigid nutations	140
10.04.3.2.3	Nutation and interior structure	142
10.04.3.3	Rotation Speed Variations	143
10.04.3.4	Polar Motion	146
10.04.3.4.1	Seasonal polar motion	146
10.04.3.4.2	Chandler wobble	147
10.04.4	Venus	148
10.04.4.1	Long-Term Spin Evolution	148
10.04.4.2	Wobble, LOD Variations, and Nutation	149
10.04.5	Mercury	150
10.04.5.1	Long-Term Spin Evolution	150
10.04.5.1.1	Spin–orbit resonance	150
10.04.5.1.2	Precession about the Cassini state and nutation	152
10.04.5.2	Librations	154
10.04.5.2.1	Free and forced libration	154
10.04.5.2.2	Forced libration and interior structure	155
10.04.5.2.3	Free libration and interior structure	157
10.04.5.3	Wobble	158
10.04.6	Summary	158
References		159

10.04.1 Introduction

The rotation of the terrestrial planets of our solar system is remarkably diverse. The Earth and Mars spin much faster than their orbital motion, whereas the rotation periods of Venus and Mercury are commensurate with their orbital periods. The two innermost planets have very small obliquities (the angle between the orbital plane and the equator of the planet) as opposed to the large obliquities of about 25° for the Earth and Mars. Mercury, the Earth, and Mars have a prograde rotation, that is, their rotational motion is in the same sense as their orbital motion; Venus, however, has a retrograde rotation. Mercury is the only solar system planet in a spin–orbit resonance: its rotation period is two-thirds of its orbital period.

Planets accumulate rotational angular momentum from the relative motion of planetesimals at collision. The angular momentum acquired in an individual impact can be large and in any direction. However, there is a high degree of cancellation for the small planetesimals and the net rotation depends on a systematic bias in the distribution of the obtained angular momentum. For a realistic distribution of planetesimal eccentricities, planets accreting from a uniform surface density disk of small planetesimals rotate progradely although much slower than the Earth and Mars (Lissauer and Kary, 1991; Lissauer *et al.*, 1997; Dones and Tremaine, 1993; Ohtsuki and Ida, 1998). A planet can acquire faster rotation in a nonuniform disk (see, e.g., Lissauer *et al.*, 2000). On the other hand, during the late stages of terrestrial planet accumulation one or a few large impacts can cause a planet to rotate very rapidly at nearly the breakup rate, so that the rotation of the planet is dominated by those random impacts. Large impacts can also determine the initial obliquity of a terrestrial planet. Due to the dominance of random phenomena and insufficient understanding of planetary accretion, definitive predictions regarding the accumulation of rotational angular momentum cannot be made for the terrestrial planets.

In this chapter, we focus on the variations in the rotation of the terrestrial planets on short timescale and review how the rotation could have evolved from the primordial rotation state. We emphasize the broad geophysical and planetological significance of rotation studies and show that they provide a direct and unique tool to explore otherwise difficult-to-obtain global properties of the terrestrial planets.

Rotation studies used to belong to the realm of celestial mechanics and the planets were usually considered to show a rigid rotational behavior. However, the internal structure and dynamics of a planet can have a large impact on its rotation. Also the atmosphere, through exchanges of angular momentum with the solid planet, can cause the rotation of a planet to change. Here, we review rotation studies from a geophysical point of point with only limited information on the celestial mechanics.

Tidal friction is a classical example of a geophysical effect on planetary rotation and the main cause of the secular braking of the rotation of the Earth. Dissipation of tidal energy in a planet delays the time of high tide, and the gravitational torque from the Sun on the tidal bulge – the tidal torque – changes the obliquity and tends to drive the planet to an equilibrium rotation rate, which is synchronous with its orbital motion for a circular orbit. Other dissipative effects, such as dissipative core–mantle interactions and thermal tides of the atmosphere can also contribute substantially to the spin evolution of the planets. Because fluid layers, such as an atmosphere and a liquid core, can rotate differentially with respect to the mantle and crust and interact with the solid, they can significantly change the rotation of the mantle and crust on short timescales. As an example, short periodic rotation variations as nutation and libration depend sensitively on the state of the core of the planet. Determination of the present core state by rotational means has profound implications for the thermal evolution and the magnetic field history of the planet, as well as for the composition and thermal state at the time of formation.

Accurate present mean rotation states have been derived from Earth-based radar observations and space missions visiting the terrestrial planets. Periodic rotation variations are more difficult to measure, and only the largest components have been obtained with limited precision. More accurate data on all aspects of rotation variations and a full exploitation of the potential of rotation studies is foreseen with improved ground-based radar observations, upcoming space missions (such as the MESSENGER and BepiColombo missions to Mercury), and plans for radio science experiments with landers on the surface of the terrestrial planets.

The outline of this chapter is as follows. In Section 10.04.2, we introduce the basic concepts and equations, which are applied to Mars, Venus, and Mercury in Sections 10.04.3, 10.04.4, and 10.04.5, respectively. We start Section 10.04.2 with the

equation of conservation of angular momentum and discuss properties of precession, nutation, polar motion, and rotation rate variations. We review in some detail the Liouville equation approach for the study of rotation variations and also briefly describe the Hamiltonian method for the case of rotation variations on long timescales. In the last part of Section 10.04.2, an overview is given of the secular effects of dissipative processes on the rotation of terrestrial planets. After an introduction on the current rotation state of Mars and its long-term variations, we present in Section 10.04.3 an overview of the nutation, length-of-day (LOD) variations, and polar motion of Mars in relation to its interior structure and atmosphere. The evolution of the rotation of Venus to its present state is reviewed in Section 10.04.4, together with its periodic rotation variations. In Section 10.04.5, we present an overview of the evolution and capture into resonance of Mercury and review its librations. In particular, we explain that accurate measurements of the librations and of the obliquity of Mercury put strong constraints on its interior structure.

10.04.2 Theoretical Foundations

10.04.2.1 Conservation of Angular Momentum

10.04.2.1.1 Precession

In this chapter, the rotation variations of the terrestrial planets are studied in a Newtonian approach. General relativistic effects are small because the squared velocities of the planetary motions are much smaller than the squared speed of light. Moreover, the ratio of the Schwarzschild radius to the radius of the planets is extremely small (for the Earth about 1.4×10^{-9}). For the rotation variations of the Earth, the general relativistic effects have been shown to be negligibly small, except for an additional contribution to precession (e.g., [Bizouard, 1995](#)). This so-called geodetic precession, though larger than the observational accuracy, can nevertheless not be separated from Newtonian precession because of uncertainties in the flattening of the Earth. Since the general relativistic effects on rotation variations are expected to be equally small for the other terrestrial planets and the observational accuracies are a few orders of magnitude larger than for the Earth, we neglect the general relativistic effects here. In this approximation, we describe the rotation variations with respect to ‘inertial reference frames’ with origin in the mass center of the planets.

The fundamental equation governing the rotation of a terrestrial planet is the classical conservation equation for angular momentum valid in an inertial reference frame:

$$\frac{d\vec{H}}{dt} = \vec{L} \quad [1]$$

Here, t is time, \vec{H} is the spin angular momentum of the planet, and \vec{L} the external torque acting on it. If no external torque is applied on the planet, eqn [1] shows that the angular momentum will not change in space. However, other solar system bodies exert a gravitational torque on the nonspherically symmetric terrestrial planets. This torque can be derived from the gravitational potential V of the planet at the position $\vec{r}(r, \theta, \lambda)$ of the mass center of a perturbing solar system body, which can generally be expressed as

$$V(\vec{r}) = -\frac{GM}{r} \sum_l \sum_{m=0}^l \left(\frac{R_e}{r} \right)^l \times [C_{lm} \cos(m\lambda) + S_{lm} \sin(m\lambda)] P_{lm}(\cos \theta) \quad [2]$$

where M is the mass of the planet, R_e the mean equatorial radius of planet, θ the co-latitude (90° – the declination), and λ the longitude (hour angle) of the perturbing body. The origin of the inertial reference frame is chosen in the mass center of the planet. The universal gravitational constant is denoted by G , and l and m are the degree and the azimuthal order of the associated Legendre function P_{lm} . Degree-one terms are zero because the origin of the coordinate system is chosen to be at the center of mass of the planet ([Heiskanen and Moritz, 1967](#)). The torque \vec{L} on the planet exerted by a solar system body with mass M^* at position \vec{r} can then be expressed as

$$\vec{L} = M^* \vec{r} \times \vec{\nabla} V \quad [3]$$

As the magnitude of the terms in eqn [2] decreases rapidly with increasing degree l for bodies far from the planet ($r \gg R_e$), only the second-degree terms are included, even in the most advanced and accurate rotation studies, except for the Earth–Moon system. In this approximation, the planet is thus considered to be a triaxial ellipsoid and the torques are proportional to the mass of the perturbing body and inversely proportional to r^3 , where r is the distance between the centers of mass of the two bodies. Therefore, only the torques of the Sun and the natural satellites orbiting the planet must be considered.

When the planet is not in a spin-orbit resonance, the external solar torque averaged over the orbital period has only a component in the orbital plane perpendicular to the rotation vector (e.g., Stacey, 1992; Chapter 3.10). Therefore, both the norm of the angular momentum and its component in the direction perpendicular to the orbital plane are constant in time. As for a spinning top, the constant (time independent) part of the gravitational torque then causes the angular momentum axis to precess about the normal to the orbit of the planet at constant obliquity. For a planet that is not in a spin-orbit resonance, the precession rate can be approximated by

$$\dot{\psi} = \alpha_p \cos \epsilon = -\frac{3}{2} \frac{n^2}{\Omega(1-e^2)^{3/2}} \mathcal{J}_2 \frac{MR^2}{C} \cos \epsilon \quad [4]$$

(see, e.g., Smart, 1953), where α_p is the precession constant, Ω is the mean rotation rate of the planet, M the total mass, R the mean radius, n the mean (orbital) motion, e the eccentricity, and $\mathcal{J}_2 \equiv -C_{20} = (C - \bar{A})/MR^2$ the degree-two zonal gravity coefficient. The principal axes of inertia of the planet are denoted by $A < B < C$, and $\bar{A} = (A + B)/2$ is the mean equatorial moment of inertia. Obliquity ϵ is the angle between the equator and the orbital plane, and longitude ψ is an orientation angle of the angular momentum axis measured from the center of the circular path of precession (we assume here that the orbit is fixed). A dotted quantity, such as $\dot{\psi}$, indicates the time-derivative of that quantity. Interesting to note is that the precession rate (eqn [4]) only depends on the degree-two, order-zero coefficient of the gravitational potential of the planet.

For rapidly rotating planets (such as the Earth and Mars), the polar flattening is essentially due to rotation and proportional to Ω^2 (e.g., Jeffreys, 1952), and so the precession rate is proportional to the rotation rate. For planets rotating so slowly that the rotational flattening is only a minor contribution to their flattening (as is the case for Mercury and Venus), the torque does not depend on the rotation rate while the spin angular momentum of the planet is linearly proportional to it. Therefore, the torque causes relatively larger changes in the angular momentum when the rotation rate deceases, such as for a spinning top, and the precession rate is inversely proportional to the rotation rate for slowly rotating planets.

The principal geophysical interest of precession is that it allows the determination of the polar moment of inertia C of the planet. Measurements of the gravitational field of the planet yield very

accurate values for differences between moments of inertia, such as \mathcal{J}_2 , but on itself cannot determine the moment of inertia (see Chapter 10.05). The moment of inertia of a planet gives insight into the mass concentration toward the center and is a major constraint for models of the interior structure of the terrestrial planets (see Chapter 10.02).

10.04.2.1.2 Nutation, polar motion, and length-of-day variations

Because terrestrial planets move on elliptical orbits, have a non-zero obliquity, and a nonspherically symmetric form, the external gravitational torque from the Sun on them is time dependent. Nutation of the angular momentum axis of a planet is the resulting small, time-dependent motion of that axis in space and can be expressed as a sum of periodic terms, whose main periods are subharmonics of the orbital period of the planet (for the Earth, see Chapter 3.10). Other planets and satellites can also exert a torque on the planet and cause additional nutational motion, but these nutations are generally negligible with respect to those related to the solar torque, except for the Earth's nutations due to the Moon.

Equation [1] shows that the change in angular momentum is entirely due to the external torque and does not depend on the interior of the planet. This suggests that nutation observations would have no geophysical interest, other than providing a value for the relative moment of inertia difference (\mathcal{J}_2/C), in addition to but less precise than precession (see eqn [4] and also eqns [64] and [65]). However, observations do not give orientation changes of the angular momentum axis but of the figure axis or the rotation axis of the mantle. In studies of periodic rotation variations, three different polar axes are considered: the rotation axis, the figure axis, and the angular momentum axis. The figure axis is defined here as the axis of greatest moment of inertia, and the rotation axis is the axis of rotation of the solid outer part of the planet, that is, mantle and crust if the (outer) core is liquid. In general, the orientation of the rotation and figure axes does depend on the interior of the planet and varies, with or without external torques, and the figure axis and the rotation axis do not coincide with the angular momentum axis. We will see that the relation between both axis and the angular momentum axes depends in a complex way on the interior, and that nutation observations give direct access to properties of the interior of a terrestrial planet.

To analyze the nutation of figure and rotation axes, it is more convenient to rewrite eqn [1] in a

reference frame attached to the rotating planet. Let $\vec{e}_1, \vec{e}_2, \vec{e}_3$ represent three orthonormal vectors along co-rotating, body-fixed axes, rotating with instantaneous angular velocity $\vec{\omega}$ with respect to an inertial reference frame. For both the inertial and the co-rotating reference frames, the center of mass of the planet is chosen as the origin O . For any vector \vec{V} , the time derivative with respect to the co-rotating reference frame is related to the time derivative with respect to the inertial reference frame as

$$\frac{d\vec{V}}{dt} \Big|_{\text{inertial}} = \frac{d\vec{V}}{dt} \Big|_{\text{co-rotating}} + \vec{\omega} \times \vec{V} \quad [5]$$

(e.g., Goldstein, 1950), and the conservation eqn [1] can be expressed as

$$\frac{d\vec{H}}{dt} + \vec{\omega} \times \vec{H} = \vec{L} \quad [6]$$

where the time derivative is with respect to the body-fixed reference frame.

The angular momentum vector can be written as an integral over the total volume of the planet, V , as

$$\vec{H} = \int_V \vec{r} \times \vec{v} \rho dV \quad [7]$$

where \vec{r} is the coordinate vector, \vec{v} the velocity, and ρ the density of the mass elements. For a rigid planet, we have

$$\vec{H} = \int_V \vec{r} \times (\vec{\omega} \times \vec{r}) \rho dV \quad [8]$$

which can, by introducing the inertia tensor \mathbf{I} with components

$$I_{ij} = \int_V [(x_1^2 + x_2^2 + x_3^2) \delta_{ij} - x_i x_j] \rho dV \quad [9]$$

be written as

$$\vec{H} = \mathbf{I} \cdot \vec{\omega} \quad [10]$$

Here, x_i are the rectangular Cartesian coordinates of the mass elements. By a suitable choice of coordinate axes, the inertia tensor can be brought to diagonal form, and the angular momentum components reduce to

$$\begin{aligned} H_1 &= A\omega_1 \\ H_2 &= B\omega_2 \\ H_3 &= C\omega_3 \end{aligned} \quad [11]$$

For principal axis rotation, it follows that the figure axis and the rotation axis coincide with the angular momentum axis for a rigid planet.

By substituting eqn [11] into eqn [6], we obtain Euler's equations for a rigid planet

$$A \frac{d\omega_1}{dt} + (C - B)\omega_2\omega_3 = L_1 \quad [12]$$

$$B \frac{d\omega_2}{dt} + (A - C)\omega_1\omega_3 = L_2 \quad [13]$$

$$C \frac{d\omega_3}{dt} + (B - A)\omega_1\omega_2 = L_3 \quad [14]$$

For rapidly spinning planets, such as the Earth and Mars, deviations from spherical symmetry are dominated by the flattening induced by rotation, which is almost symmetric about the x_3 -axis. Therefore, the approximation $A = B$ is often used. For the Earth, for example, the relative difference between A and B is more than 100 times smaller than the relative difference between A and C . In the absence of external torques \vec{L} , the solution of Euler's equations can then be written as

$$\begin{aligned} \omega_1 &= a \cos(\sigma_E t + \phi) \\ \omega_2 &= a \sin(\sigma_E t + \phi) \\ \omega_3 &= \Omega \end{aligned} \quad [15]$$

Given an initial deviation of the rotation axis from the figure axis, the rotation axis rotates, or wobbles, about the figure axis with the Euler frequency

$$\sigma_E = \frac{C - A}{A} \Omega \quad [16]$$

This Euler wobble is a manifestation of the conservation of angular momentum when the figure axis and the rotation axis do not coincide. Solution [15] is an example of wobble or polar motion, which in general describes the orientation variations of the rotation axis with respect to a body-fixed polar axis (for more precise definitions of polar motion and wobble, see Chapter 3.10). Seasonal polar motion is mainly caused by motion in fluid planetary layers, such as an atmosphere and oceans. These motions can also cause periodic variations in the ω_3 -component of the rotation vector, or changes in the rotation speed. As these variations imply changes in the length of the day, they are, in the geophysical literature, mostly referred to as length-of-day variations (ΔLOD).

10.04.2.1.3 Kinematical relation between polar motion and nutation

To describe rotation variations in space, the body-fixed reference frame must be related to an inertial reference frame, which we here take to have coordinates x'_1, x'_2, x'_3 and an origin at the mass center of the

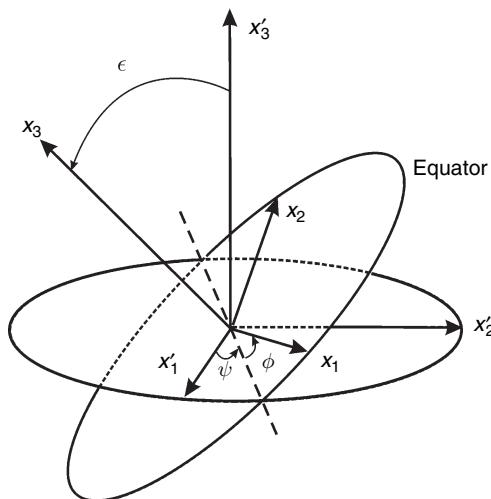


Figure 1 Euler angles.

planet. The transformation is usually described with three rotations over three Euler angles (see **Figure 1**). We introduce the Euler angle ψ from the x'_1 -axis to the intersecting line (ascending node line) between the (x'_1, x'_2) -plane of the inertial reference frame and the equator of the body-fixed reference frame, the Euler angle ϵ from the inertial x'_3 -axis to the normal to the planet's equator (x_3 -axis), and the Euler angle ϕ from the node line to the x_1 -axis. If the orbital plane is assumed fixed and coincides with the (x'_1, x'_2) -plane ψ is the precession angle, ϵ the obliquity, and ϕ the rotation angle. The time derivatives of the Euler angles represent combinations of rotations of the three body-fixed axes with respect to the inertial frame. The relation between the components (in the body frame) of the rotation vector and the time derivatives (in inertial space) of the Euler angles is given by Euler's kinematical equations (e.g., Goldstein, 1950; Moritz and Mueller, 1987)

$$\begin{aligned}\omega_1 &= -\dot{\epsilon} \cos \phi - \dot{\psi} \sin \epsilon \sin \phi \\ \omega_2 &= \dot{\epsilon} \sin \phi - \dot{\psi} \sin \epsilon \cos \phi \\ \omega_3 &= \dot{\phi} + \dot{\psi} \cos \epsilon\end{aligned}\quad [17]$$

Because the figure axis is defined in terms of the planetary mass-elements that rotate about the rotation axis, a variation in the rotation axis with respect to an inertial reference frame also leads to a variation in the figure axis with respect to the inertial reference frame. Moreover, these variations are not the same, and therefore any periodic variation in the orientation of the rotation or figure axis in space

(nutation) is accompanied by a periodic variation of the rotation axis with respect to the figure axis (polar motion or wobble), and vice versa. Consider, for example, the Euler wobble (eqn [15]). During the wobble, both the figure and the rotation axis have to change their orientation in inertial space; otherwise, the planet's spin angular momentum (which is defined in terms of the inertia and rotation of the planet, see eqn [10]) cannot be conserved. Mathematically, from the first two of Euler's kinematical eqns [17], we have

$$\dot{\epsilon} + i \sin \epsilon \dot{\psi} = -\tilde{\omega} e^{i\phi} \quad [18]$$

where $\tilde{\omega} = \omega_1 + i\omega_2$ is the complex equatorial component of the rotation. If the x'_3 -axis is chosen normal to the ecliptic of date of the planet and the x'_1 -axis points to the ascending node, $\dot{\epsilon}$ and $\dot{\psi}$ represent changes in obliquity and in longitude (nutation). Equation [18] relates wobble ($\tilde{\omega}/\Omega$) to the time-derivative of the nutation of the figure axis ($\dot{\epsilon} + i \sin \epsilon \dot{\psi}$). When the time dependence of ϕ is approximated by Ωt , periodic changes in the rotation vector with frequency σ in the co-rotating reference frame (wobble) are seen to correspond to changes in longitude and obliquity (nutation) with frequency $\sigma' = \sigma + \Omega$ in the inertial reference frame.

By integrating eqn [18] with respect to time, it also follows that the amplitude of nutation of the figure axis differs by a factor Ω/σ' from the amplitude of the wobble ($\tilde{\omega}/\Omega$) of the rotation axis about the planet's figure axis. The motion of the rotation axis in space is the sum of the motion of the rotation axis about the figure axis (in Euler angles: $\delta\epsilon + i \sin \epsilon \delta\psi = -i\tilde{\omega} \exp(i\Omega t)/\Omega$; see, e.g., Moritz and Mueller, 1987) and the motion of the figure axis in space. It then follows that the ratio of the amplitude of the wobble and the amplitude of nutation of the rotation axis is equal to σ'/σ . For the rapidly rotating planets Earth and Mars, the rotation period is much shorter than the orbital period. The main periods of nutation are related to the orbital motion and are therefore long compared to the rotation period (σ'/Ω small). Wobble periods are related to the seasons (and thus also to the orbital motion) and are therefore also long (σ/Ω small). The main nutations are therefore associated with very small wobble of diurnal period in the co-rotating reference frame, and large polar motion is associated with very small nutation of the rotation axis (i.e., of diurnal period in space). When

observations have limited time resolution, averaged orientations can only be obtained, and the small diurnal contributions to polar motion and nutation are neglected. For the Earth, the convention is to introduce an intermediate polar axis that only has low-frequency nutational motion in space, and to define polar motion as the motion of that axis in the terrestrial reference frame outside the retrograde diurnal frequency band (e.g., Chapter 3.10). Each phenomenon can then be studied separately. Polar motion has been described before as the motion of the rotation axis about the figure axis, but actually we see that the rotation axis remains nearly fixed with respect to inertial space. It is the figure axis, or the planet itself, that wobbles about the rotation axis.

10.04.2.2 Liouville Equations

10.04.2.2.1 Excitation

Both external and internal processes can excite periodic rotation variations of a terrestrial planet. We have sufficient knowledge about the external gravitational causes, but other processes are more difficult to model. For example, the seasonal global CO₂ exchange between the Martian atmosphere and the polar caps causes variations in the LOD of Mars, but is not precisely known. By studying rotational variations of a terrestrial planet, more can thus be learned about the excitation processes. Moreover, as the rotational response depends on the planet's structure and composition, also insight into the planetary interior can be obtained. This is particularly so for the rotational variations due to well-known external gravitational causes, such as for the nutations of Mars and the libration of Mercury.

10.04.2.2.2 Basic equations

Basic to most geophysical studies of the rotation of the Earth are the Liouville equations, which describe the change in angular momentum \vec{H} of the deformable planet due to an applied torque (see Chapter 3.10). These equations are also an ideal tool for investigating the rotation of the other terrestrial planets. We use a co-rotating reference frame as before. As a terrestrial planet is not rigid, we consider the three coordinate axes to be along the mean axes of principle moments of inertia of the mantle and to rotate with instantaneous angular velocity $\vec{\omega}$ with respect to an inertial reference frame. In equilibrium, the planet is assumed to rotate uniformly with

angular velocity $\vec{\omega} = \Omega \vec{e}_3$. Due to internal and external excitations, the rotation velocity of the mantle

$$\vec{\omega} = \Omega \begin{pmatrix} m_1 \\ m_2 \\ 1 + m_3 \end{pmatrix} \quad [19]$$

deviates from the equilibrium rotation velocity aligned with the polar axis. The small deviations m_1 and m_2 describe wobble, and m_3 gives LOD variations.

In the reference frame rotating with the instantaneous rotation vector of the mantle, the change in angular momentum of the planet is governed by eqn [6]. The angular momentum (eqn [7]) can be expressed as

$$\vec{H} = \mathbf{I} \cdot \vec{\omega} + \vec{b} \quad [20]$$

where

$$\mathbf{I} = \begin{pmatrix} A & 0 & 0 \\ 0 & B & 0 \\ 0 & 0 & C \end{pmatrix} + (c_{ij}) \quad [21]$$

Mass redistribution and deformation of the planet cause small increments in inertia c_{ij} , and the small relative angular momentum \vec{b} is due to relative motion with respect to the co-rotating frame in the fluid parts of the planet (liquid core, oceans, and atmosphere). We assume here that there is no rotation in the deformation field of the mantle ($\vec{b}^{\text{mantle}} = 0$) with respect to the reference frame of the planet. The axes are then called mantle Tisserand axes (Munk and MacDonald, 1960; Moritz and Mueller, 1987). Equation [20] shows that part of the angular momentum is absorbed in relative motion, such as relative motion of a liquid core, and in perturbations of the moment of inertia. Rotation variations for a deformable planet are therefore different from those of a rigid planet. Deformations tend to decrease the rotation variations. A rotationally decoupled liquid core, on the other hand, may lead to larger mantle rotation variations because the mantle has a smaller moment of inertia than the total planet.

By using eqns [19]–[21], the three components of the Liouville equations for the whole planet can be expressed as

$$A\Omega \dot{m}_1 + \Omega^2(C - B)m_2 = L_1 - \Omega \dot{c}_{13} + \Omega^2 c_{23} - \dot{b}_1 + \Omega b_2 \quad [22]$$

$$B\Omega \dot{m}_2 - \Omega^2(C - A)m_1 = L_2 - \Omega \dot{c}_{23} - \Omega^2 c_{13} - \dot{b}_2 - \Omega b_1 \quad [23]$$

$$C\Omega \dot{m}_3 = L_3 - \Omega \dot{c}_{33} - \dot{b}_3 \quad [24]$$

These equations are linearized in m_i and c_{ij} , as these quantities are very small compared to their leading order terms in the equations. A particular consequence of this approximation is that there is no term in $B - A$ in eqn [24], because that term is of second order in the rotation variations (e.g., Hinderer *et al.*, 1982). Furthermore, the first two equations decouple from the third equation. Therefore, wobble and LOD variations are decoupled and can be calculated separately. The symmetry between eqn [24] on the one hand and eqns [22] and [23] on the other hand is broken because the planet's mean rotation introduces a preferred direction, which we have chosen to be along the x_3 -axis.

The right-hand members of the eqns [22]–[24] describe the excitations: (1) external torques, (2) mass redistribution, and (3) relative motion. Of these, the external torque can be determined most precisely. The mass redistribution and relative motion on the other hand not only depend on the excitation process but also on the rotation variations, as we will see below. The right-hand members of eqns [22]–[24] divided by $\Omega(C - A)$ are called the excitation functions (Munk and MacDonald, 1960).

For rapidly rotating planets, the equatorial flattening (due to nonhydrostatic effects) is much smaller than the polar flattening (mainly due to the fast rotation), so the approximation $A = B$ can be used in the Liouville equations. By introducing the complex notation

$$\begin{aligned}\tilde{m} &= m_1 + i m_2 \\ \tilde{c} &= c_{13} + i c_{23} \\ \tilde{b} &= b_1 + i b_2\end{aligned}\quad [25]$$

wobble can be described by the complex sum of the first two components of the Liouville equations [22] and [23]:

$$A\Omega\dot{\tilde{m}} - i(C - A)\Omega^2\tilde{m} = -\Omega\dot{\tilde{c}} - i\Omega^2\tilde{c} + \dot{\tilde{b}} - i\Omega\tilde{b} \quad [26]$$

If the planet's core is liquid, Hough (1895, see also Poincaré, 1910) showed with mathematical rigour that a change in the mantle rotation leads to a change in the core rotation that is of the same order of magnitude as the mantle rotation change, and an additional irrotational core flow, which is one order in the core flattening smaller (the total core flow is usually referred to as a Poincaré flow). This result was derived for an incompressible, inviscid and homogeneous core, but further studies showed it to be a valid approximation when these conditions are relaxed (see, e.g., Stewartson and Roberts, 1963;

Roberts and Stewartson, 1965; Busse, 1968, for viscosity; and Smith, 1977 for a numerical investigation of the problem). Three additional quantities m_i^f , describing the relative rotation of the fluid core with respect to the mantle, have then to be determined. We define these by

$$\vec{\omega}^f = \Omega \begin{pmatrix} m_1^f \\ m_2^f \\ m_3^f \end{pmatrix} \quad [27]$$

The total rotation vector of the core is then $\vec{\omega} + \vec{\omega}^f$. It is assumed here that the core co-rotates with the mantle in the equilibrium state. The differential rotations of the core are determined by three additional Liouville equations, which express the rate of change of angular momentum of the core as a result of the torque \vec{L}^f exerted on the core in the same mantle co-rotating reference frame as before:

$$\frac{d\vec{H}^f}{dt} + \vec{\omega} \times \vec{H}^f = \vec{L}^f \quad [28]$$

where

$$\vec{H}^f = \mathbf{I}^f \cdot \vec{\omega} + \vec{b}^f \quad [29]$$

$$\begin{aligned}&= \Omega \left[\begin{pmatrix} A^f & 0 & 0 \\ 0 & B^f & 0 \\ 0 & 0 & C^f \end{pmatrix} + \begin{pmatrix} c_{ij}^f \\ c_{ij}^f \\ c_{ij}^f \end{pmatrix} \right] \\ &\quad \times \begin{pmatrix} m_1 \\ m_2 \\ 1 + m_3 \end{pmatrix} + \begin{pmatrix} A^f m_1^f \\ B^f m_2^f \\ C^f m_3^f \end{pmatrix} \quad [30]\end{aligned}$$

where c_{ij}^f are small increments in the inertia of the core. The three components of the Liouville equations for the liquid core can then be written as

$$\begin{aligned}A^f \Omega \left(\dot{m}_1 + \dot{m}_1^f \right) - B^f \Omega^2 m_2^f + \Omega^2 (C^f - B^f) m_2 \\ + \Omega \dot{c}_{13}^f - \Omega^2 c_{23}^f = L_1^f\end{aligned}\quad [31]$$

$$\begin{aligned}B^f \Omega \left(\dot{m}_2 + \dot{m}_2^f \right) + A^f \Omega^2 m_1^f - \Omega^2 (C^f - A^f) m_1 \\ + \Omega \dot{c}_{23}^f + \Omega^2 c_{13}^f = L_2^f\end{aligned}\quad [32]$$

$$C^f \Omega \left(\dot{m}_3 + \dot{m}_3^f \right) + \dot{c}_{33}^f \Omega = L_3^f \quad [33]$$

where only terms linear in m_i , m_i^f , c_{ij} , and c_{ij}^f were retained. As for the Liouville equations for the whole planet, the rotation rate variations are decoupled from the orientation variations, and there is no term

in $B^f - A^f$ in eqn [33]. To obtain the solutions for wobble (m_x and m_y) and LOD variations (m_z) of a planet with a liquid core, the deformation induced incremental inertia terms (c_{13} , c_{23} , c_{13}^f , c_{23}^f , c_{33} , c_{33}^f), the relative angular momentum (b_1 , b_2 , b_3), and the torques (L_1 , L_2 , L_3 , L_1^f , L_2^f , and L_3^f) have to be determined from the excitation processes considered. For every frequency σ , the associated nutations of the rotation axis have an amplitude differing by a factor σ/σ' as follows from integrating Euler's kinematical equation [18] with respect to time.

10.04.2.2.3 Core–mantle coupling

The torque on the fluid core is the sum of the external gravitational torque on the permanent core figure due to the Sun and other solar system bodies and the torque due to the mantle. The latter torque includes viscous coupling between mantle and core, electromagnetic coupling, gravitational coupling, pressure coupling due to the ellipsoidal core form, and pressure coupling due to topography on the core–mantle boundary (CMB). By assuming the dissipative torques (viscous and electromagnetic) to be negligibly small for relatively short-period rotation variations (but see Section 2.6 for their effect on the long-term rotation) and by neglecting CMB topography other than the ellipticity, the torque on the fluid core is often written as the sum of the external gravitational torque (\vec{L}_{ext}^f) and the so-called inertial torque (\vec{N}), which is the sum of the gravitational and pressure torque. By using the Navier–Stokes equation for the fluid core, the inertial torque can be calculated by assuming a Poincaré flow in the fluid core. Correct up to the first order in the flattenings and the rotation variations, we have

$$\begin{aligned} N_1 &= -L_{\text{ext},1}^f + \Omega^2(C^f - B^f)(m_2 + m_2^f) - \Omega^2 c_{23}^f \\ N_2 &= -L_{\text{ext},2}^f + \Omega^2(A^f - C^f)(m_1 + m_2^f) + \Omega^2 c_{13}^f \\ N_3 &= -L_{\text{ext},3}^f \end{aligned} \quad [34]$$

(e.g., Rochester, 1976; Sasao *et al.*, 1977, 1980; Hinderer *et al.*, 1982; and Van Hoolst and Dehant, 2002 for the triaxial case). By substituting these expressions into Liouville equations [31]–[33] for the core, we have

$$\frac{d\vec{H}^f}{dt} - \vec{\omega}^f \times \vec{H}^f = 0 \quad [35]$$

(Sasao *et al.*, 1980; Mathews *et al.*, 1991). From expression [34] for the third component of the inertial torque, it follows that $L_3^f = 0$ and

$$C^f \Omega \left(\dot{m}_3 + \dot{m}_3^f \right) + \dot{c}_{33}^f \Omega = 0 \quad [36]$$

If deformation is neglected, there is no net change in the third component of the angular momentum of the core in the lowest-order approximation in the rotation variations and the flattenings: the external torque is balanced by the core–mantle coupling mechanisms for the third component. For an incompressible, inviscid, and homogeneous core, Hough (1895) already demonstrated that the additional core rotation with respect to the equilibrium rotation is purely equatorial, so that the rotation rate of the core remains constant. There is no complete torque balance for the equatorial components. Again, this symmetry-breaking between the axial and equatorial components is due to the mean rotation of the planet, which introduces a preferred direction in the problem.

10.04.2.2.4 Mass redistribution

We now discuss the problem of determining the incremental inertia terms. In particular, we consider the interesting case of excitation by an atmosphere as a typical example that includes excitation due to both relative motion and mass redistribution. By setting the external torque equal to zero and introducing the equatorial angular momentum function $\tilde{\chi} = (\Omega\tilde{c} + \tilde{b})/((C-A)\Omega)$ and the axial angular momentum function $\chi_3 = (\Omega c_{33} + b_3)/C\Omega$, the Liouville equations [26] and [24] can be re-expressed as

$$\tilde{m} + \frac{i}{\sigma_E} \frac{d\tilde{m}}{dt} = \tilde{\chi} - \frac{i}{\Omega} \frac{d\tilde{\chi}}{dt} \quad [37]$$

$$\dot{m}_3 = -\dot{\chi}_3 \quad [38]$$

The contribution of the atmosphere to the angular momentum functions can be separated in a term due to the wind (\tilde{b} and b_3) and a term due to the atmospheric mass (\tilde{c} and c_{33}). Explicit expressions for these ‘wind’ and ‘matter’ atmospheric angular momentum functions in terms of wind velocity and atmospheric pressure can be found in, for example, Barnes *et al.* (1983). For a rigid planet, only these atmospheric wind angular momentum functions, $\tilde{\chi}_A^W$ and $\chi_{3,A}^W$, and atmospheric matter angular momentum functions, $\tilde{\chi}_A^M$ and $\chi_{3,A}^M$ have to be considered in the Liouville equations [37] and [38]. The superscript

A denotes that only the direct contribution from the atmosphere (and seasonal polar caps) is considered, and superscripts M and W indicate matter and wind terms, respectively.

For an entirely solid but nonrigid planet, contribution to the angular momentum functions from the incremental inertia terms $\tilde{c} = c_{13} + ic_{23}$ and c_{33} due to the planet's response have to be determined additionally. Two effects contribute to these inertia terms. First, when the rotation vector of the planet changes, the centrifugal potential changes and induces a slight mass redistribution of the planet. The associated changes in the inertia terms (indicated by superscript 'centr') can be determined for any planet from MacCullagh's theorem

$$\tilde{c}^{\text{centr}} = k_2 \frac{\Omega^2 R^5}{3G} \tilde{m} \quad [39]$$

$$c_{33}^{\text{centr}} = k_2 \frac{4R^5 \Omega^2}{9G} m_3 \quad [40]$$

where k_2 is the potential Love number of degree 2 of the planet (Munk and MacDonald, 1960; Moritz and Mueller, 1987). These contributions to the inertia terms are independent of the excitation source.

A second contribution to the incremental inertia terms \tilde{c} and c_{33} is due to the atmosphere (and ice cap) loading. The loading induces a deformation and a corresponding change in the mass distribution of the planet. The associated change in the planetary gravitational potential at the planet's surface is most conveniently expressed as k'_2 times the potential of the load, where k'_2 is the degree-two load Love number (Munk and MacDonald, 1960). Since the incremental moments of inertia associated with the change in the gravitational potential due to the atmospheric loading are linearly proportional to the change in the gravitational potential according to MacCullagh's theorem, it follows that the contribution of the loading effect is $k'_2 \chi_A^M$ to the equatorial angular momentum function and $k'_2 \chi_{3,A}^M$ to the axial angular momentum function.

10.04.2.2.5 Periodic rotation variations

By taking the mass redistribution into account and solving eqns [37] and [38] at a given frequency σ , wobble of a deformable solid planet can be expressed as

$$\tilde{m}(\sigma) = -\frac{\alpha(\sigma + \Omega)}{(\sigma - \sigma_{\text{CW}})} [(1 + k'_2) \tilde{\chi}_A^M(\sigma) + \tilde{\chi}_A^W(\sigma)] \quad [41]$$

(e.g., Moritz and Mueller, 1987), and relative rotation rate variations are given by

$$\begin{aligned} m_3(\sigma) &= -[(1 + k'_2) \chi_{3,A}^M(\sigma) + \chi_{3,A}^W(\sigma)] \\ &\times \left(1 + k_2 \frac{4R^5 \Omega^2}{9GC}\right)^{-1} \end{aligned} \quad [42]$$

Here,

$$\alpha = \frac{C - (A + B)/2}{(A + B)/2} \quad [43]$$

is the polar flattening. The only differences with respect to a solid planet are the term in the loading Love number k'_2 , and the change from the Euler frequency σ_E to the frequency σ_{CW} of the Chandler wobble (CW), which is the extension of the Euler wobble to a deformable planet (see also eqn [46] below). The difference between the Chandler wobble frequency and the Euler frequency is due to the deformation of the planet as a result of the change in the centrifugal potential associated with the wobble.

For planets with a liquid core, the rotation rate variations differ only by a factor C/C^m with respect to a solid planet:

$$m_3(\sigma) = -\frac{C}{C^m} [(1 + k'_2) \chi_A^M(\sigma) + \chi_A^W(\sigma)] \quad [44]$$

when the small inertia terms due to the centrifugal force are neglected. These terms describe the additional variation in rotation rate due to the changes in the polar moment of inertia resulting from the rotation rate variations, and are a second-order effect.

For polar motion of planets with a liquid core, the incremental inertia terms $\tilde{c}^f = c_{13}^f + ic_{23}^f$ and c_{33}^f do have to be determined. The main reason is that changes in the inertia tensor are compared to differences in the moments of inertia for polar motion (see eqns [22], [23], [31], [32]), and for rotation rate variations they are compared to the polar moment of inertia C itself (see eqns [24] and [33]), which is several orders of magnitude larger for the terrestrial planets. The incremental core inertia terms can be expressed in terms of compliances or generalized Love numbers in the case of a homogeneous core (Sasao *et al.*, 1977, 1980; Hinderer *et al.*, 1982; Dehant *et al.*, 1993; see Chapter 3.10). By solving the Liouville equations for the entire planet and the liquid core, wobble can then be written as

$$\begin{aligned} \tilde{m}(\sigma) &= -\frac{\alpha A(\sigma + \Omega)}{A^m(\sigma - \sigma_{\text{CW}})} \left(1 + \frac{A^f}{A} \frac{\sigma'}{\sigma' - \sigma'_{\text{FCN}}}\right) \\ &\times [(1 + k'_2) \tilde{\chi}_A^M(\sigma) + \tilde{\chi}_A^W(\sigma)] \end{aligned} \quad [45]$$

(Sasao *et al.*, 1980; Defraigne *et al.*, 2000). Here, σ_{CW} and σ'_{FCN} are the frequencies of the Chandler wobble and the Free Core Nutation (FCN). These normal modes can be determined by solving the Liouville equations in the case of no internal or external excitation. The CW consists essentially of a rigid rotation of the mantle about the mantle rotation axis. The FCN is mainly a rigid rotation of the fluid core about the mean mantle rotation axis and only exists when the core is liquid. As they are defined in the absence of external excitation, the normal modes conserve the total spin angular momentum of the planet. Both the CW and the FCN have their origin in the flattening of the planet. The frequencies of the CW and the FCN can, correct up to the first order in the flattenings, be expressed as (Sasao *et al.*, 1980; Yoder and Standish, 1997; Van Hoolst and Dehant, 2002; Chapter 3.10)

$$\sigma_{\text{CW}} = \Omega \frac{A}{A^m} \sqrt{\left(1 - \frac{k_2}{\kappa}\right)^2 \alpha^2 - \frac{\beta^2}{4}} \quad [46]$$

$$\sigma'_{\text{FCN}} = -\Omega(\alpha_f - \beta_d) A / A_m \quad [47]$$

where $\kappa = 3\mathcal{J}_2/q_0$ the is the secular Love number, α_f the dynamical polar flattening at the CMB, and β_d a compliance that describes the deformations due to the centrifugal effect of the differential wobble of the core. We also used the equatorial flattening

$$\beta = \frac{B - A}{A} \quad [48]$$

and the ratio of the surface centrifugal acceleration to the surface gravity

$$q_0 = \frac{\Omega^2 R^3}{GM} \quad [49]$$

As the flattenings are small for terrestrial planets, both these periods are long compared to the rotation period of the planet. Note, however, that the CW period is long in a co-rotating reference frame, and that the FCN has a long period with respect to an inertial reference frame. Equations [46] and [47] show that deformation ($k_2 > 0$ and $\beta_d > 0$) lengthens the CW and FCN periods and that a liquid core decreases the CW period.

Solutions [41] and [45] for wobble of a solid planet and a planet with a liquid core show resonances at the frequencies of the CW and the FCN. These normal modes of the planet are characterized by a predominant rotational motion and are sometimes called rotational normal modes. In general, the rotational response of a planet to an external forcing can be expressed as a linear combination of the planet's normal eigenfunctions, which form a complete set (Wahr, 1981a, 1981b).

Although the largest, and most familiar, class of normal modes is the set of seismic modes, the rotational normal modes are more important for the description and understanding of planetary rotation variations. The seismic modes have very small rotation in their displacement field, and therefore cannot appreciably affect the rotation of a planet. They also have much smaller periods than the periods at which rotation variations are excited. For the Earth, the gravest mode is the so-called football mode, with a period of 53.9 min (Dahlen and Tromp, 1998). All other modes have smaller periods, typically of the order of minutes or tens of minutes. The three other terrestrial planets are smaller and less dense, and, being of similar structure, their seismic periods are smaller. As an example, for Mars the football mode has a period of about 30 min (Van Hoolst *et al.*, 2000a; Gudkova and Zharkov, 2004). As the rotational forcing frequencies are far from the frequencies of the seismic normal modes, there are no resonance effects with these modes.

Four rotational normal modes mainly related to orientation changes can be distinguished for terrestrial planets with at most three different internal layers that can show relative rotation (solid mantle + crust, liquid outer core, and solid inner core): the Chandler wobble (CW), the Free Core Nutation (FCN), the Inner Core Wobble (ICW), and the Free Inner Core Nutation (FICN). The ICW is analogous to the CW and describes the relative rotation of the figure axis of the inner core about the mantle rotation axis. The FICN represents the relative rotation of the inner core about the mantle rotation axis. The ICW has a long period in a co-rotating frame and the FICN a long period in an inertial frame. Extended Liouville theories for terrestrial planets with a solid elastic mantle, a liquid outer core, and an elastic inner core show that the solution for polar motion shows resonances at the eigenfrequencies of these four rotational normal modes (see, e.g., Mathews *et al.*, 1991, 2002; and Dehant *et al.*, 1993, 2005). An important consequence of these resonances is that for frequencies close to the resonant frequencies, rotation variations can be amplified. Detection of the amplification yields information on the normal-mode frequencies, and thus on the internal properties on which these frequencies depend (see, e.g., eqns [46] and [47]).

10.04.2.3 Hamiltonian Equations and Long-Term Evolution

In astronomical studies of planetary rotation variations in which the planets are considered to be rigid

(as opposed to geophysical studies), Hamiltonian and Lagrangian approaches are customarily used (Smart, 1953; Kinoshita, 1977). These methods have also been extended to include effects of deformation and core–mantle coupling for the nutations of the Earth (see, e.g., Getino and Ferrández, 2000) but have not yet reached that level of precision for the short-term variations in the rotation of the other terrestrial planets. On the other hand, for studies of the long-term evolution of planetary spins, Hamiltonian and Lagrangian equations are often preferred to torque equations. Here, we briefly introduce the Hamiltonian approach to planetary rotation in the frame of long-term changes of rotation.

We assume, as is often done, that the terrestrial planets rotate about their greatest principal inertia axis and that the angular momentum axis also coincides with that axis, as would be the case for a rigid planet with principle axis rotation. Besides neglecting polar motion, nutation is also assumed to be zero. These assumptions are justified since polar motion and nutation are very small compared to the long-term changes (see Sections 10.04.3, 10.04.4, and 10.04.5) and average to zero over a sufficiently long time. Moreover, large wobble at any moment in the history of a planet will be damped in a relatively short time compared to the age of the solar system (Peale, 1976a; Yoder and Ward, 1979).

Basic to the Hamiltonian and Lagrangian methods is the development of expressions for the kinetic and potential energy of the rotating planet interacting with other solar system bodies. The rotational kinetic energy T of a planet supposed to be rigid is given by

$$T = \frac{1}{2}(A\omega_1^2 + B\omega_2^2 + C\omega_3^2) \quad [50]$$

when planetary effects on the orbit are neglected (see, e.g., Goldstein, 1950). The components of the angular velocity $\vec{\omega}$ are with respect to the principal axes of inertia of the planet. The potential energy of an isolated rigid body does (almost) not change with time and can therefore be neglected. The only potential energy to be included then is that due to the gravitational interaction with the Sun, the other planets, and satellites. By retaining only the degree-two terms of eqn [2], the potential energy U for an interacting body with mass M^* can be expressed as

$$U = -\frac{GMM^*}{r} - \frac{GMM^*}{2r^3}(A + B + C - 3\mathcal{I}) \quad [51]$$

where \mathcal{I} is the moment of the inertia of the planet with respect to the direction from the planet's mass center to the perturbing body.

In the Lagrangian equations,

$$\frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{q}_n}\right) - \frac{\partial \mathcal{L}}{\partial q_n} = 0 \quad [52]$$

where $\mathcal{L} = T - U$ is the Lagrangian, the Euler angles are taken as the generalized coordinates q_n . By using eqn [17], the kinetic energy can readily be expressed in terms of the time derivatives of the Euler angles. For Hamilton's equations, the Euler angles and their adjoint moments could be used. However, these moments have no intuitive physical meaning, and often Andoyer variables are preferred (Andoyer, 1923; Kinoshita, 1977). For the case of principal axis rotation, the problem is reduced from six variables to four variables (Kinoshita, 1977). The two Andoyer variables are (1) the angular momentum amplitude $H = C\omega$ and (2) the angular momentum component normal to the orbital plane of date, $X = H\cos\epsilon$, and the two conjugate Andoyer variables are (1) the hour angle θ between the vernal equinox of date and a fixed direction in the equator (e.g., the axis of minimum moment of inertia), and (2) the precession ψ (in general taking into account the motion of the orbit: see, e.g., Kinoshita, 1977). Besides on the hour angle, the Hamiltonian also depends on a second fast variable describing the revolution of the planet around the Sun. For studies of the long-term evolution of the planet, one averages the Hamiltonian over these fast angles. For a rigid planet with principal axis rotation and without planetary perturbations and spin-orbit resonances, the resulting averaged Hamiltonian can be expressed as

$$\mathcal{H}_{\text{averaged}} = \frac{H^2}{2C} - \frac{\alpha_p}{2} \frac{X^2}{H} \quad [53]$$

(see, e.g., Néron de Surgy and Laskar, 1997), where α_p is the precession constant (eqn [4]), and the Hamiltonian equations yield

$$\begin{aligned} \frac{dH}{dt} &= 0 \\ \frac{dX}{dt} &= 0 \\ \frac{d\theta}{dt} &= \frac{H}{C} + \frac{\alpha_p}{2} \frac{X^2}{H^2} \\ \frac{d\psi}{dt} &= \alpha_p \frac{X}{H} \end{aligned} \quad [54]$$

Within the approximations made, the rotation speed of the planet, the obliquity, and the precession rate are constant in time.

In the case of spin–orbit resonances, that is, when the ratio of the rotation speed ω and the mean motion n approximates a rational number $p = \text{integer}/2$, combinations $\theta - pM$ of the two fast variables (M is the mean anomaly) appearing in the Hamiltonian are not fastly variable, and cannot be eliminated by the averaging procedure. The Andoyer variables H and X , or equivalently the rotation speed and the obliquity, will then become time dependent, with their time derivatives being proportional to $\sin 2(\theta - pM)$, and spin–orbit resonances also contribute a term to the precession (see, e.g., Correia *et al.*, 2003). For Mercury, the resulting changes in rotation speed (librations) and obliquity variations, will be studied in Section 10.04.5.

Additional effects can be taken into account by adding the corresponding energy to the Hamiltonian. Effects related to dissipative processes will be studied in Section 10.04.2.4. Here, we briefly consider the inclusion of planetary perturbations. Because the ecliptic of date, with respect to which the Andoyer variables are defined, changes its orientation due to planetary gravitational effects, additional terms have to be added to the Hamiltonian eqns [54] for X and ψ (e.g., Peale, 1973; Néron de Surgy and Laskar, 1997). The direct planetary gravitational torque on a planet is much smaller than the indirect effects of the other planets on the orbit. As mentioned above, the direct torque is usually neglected and so H remains constant. Planetary effects alone therefore do not affect the rotation rate of a planet, but do change the obliquity and precession.

If the spin axis of a terrestrial planet were fixed in inertial space, obliquity variations would be entirely due to the planetary induced variations in the orientation of the orbital plane (called orbital precession). However, the solar torque (and also the lunar torque for the Earth) on the flattened planet causes the planet to precess about the orbit normal. The resulting motion of the spin axis with respect to the orbit normal depends critically on the ratio of the precession period to the period of the orbital precession. For a very small ratio, the spin axis precession is about the slowly moving orbit normal and tracks that motion efficiently such that the obliquity is nearly constant. On the other hand, if the precession of the spin axis is much slower than the precession of the orbit, the spin axis will precess about the averaged orbit normal (see, e.g., Ward, 1992). If the ratio

is about one (as is now the case for Mars), we have a resonant situation, and the obliquity variations can be large and chaotic (e.g., Ward, 1974; Touma and Wisdom, 1993). Laskar and Robutel (1993) have shown that the four terrestrial planets could have experienced large chaotic variations in obliquity during some period(s) of their history, and that their obliquities cannot be considered as primordial.

10.04.2.4 Dissipative Effects as Causes of Secular Rotation Variations

10.04.2.4.1 Gravitational tides

Dissipation of tidal energy in the Earth is the principal cause of the secular increase in the LOD (Munk and MacDonald 1960). Deviations from perfect elasticity of the planetary material delay the response of the planet to an applied tidal torque and cause the tidal bulge to be slightly misaligned with respect to the direction to the body raising the tides. The misalignment results in a small torque on the planet that alters the rotation of the planet. The theory of tides and the influence on the rotation and orbits of planets and satellites have been described in many books and review articles, and will not be repeated here in detail (see, e.g., Darwin, 1908; Munk and MacDonald, 1960; Kaula, 1964; MacDonald, 1964; Goldreich and Soter, 1966; Melchior, 1983; Lambeck, 1980; Murray and Dermott, 1999). An important property is that the main part of the tidal potential is proportional to r^{-3} (see the degree-two part of eqn [2]), where r is the distance between the planet and the tide-raising body. Therefore, the tidal bulge is also proportional to r^{-3} and the tidal torque is proportional to r^{-6} . As for the Earth, the solar tidal torque slows down the rotation of a rapidly rotating terrestrial planet ($\omega > n$ or $\omega < -n$) since then the tidal bulge leads the planet–Sun direction when measured in the sense of rotation of the planet.

Because of the strong dependence on the distance to the Sun, tidal braking due to the Sun is much more efficient for Mercury and Venus than for the Earth and Mars and could be the cause of the slow rotation of Mercury and Venus compared to the fast rotation of the Earth and Mars. From eqn [1] and a tidal torque of magnitude

$$L_T = -\frac{3k_2 GM_{\odot}^2 R^5}{2a^6} \sin 2\delta \quad [55]$$

for a circular orbit and zero obliquity (e.g., MacDonald, 1964; Murray and Dermott, 1999), the

rate of decrease of rotation velocity can be approximated by

$$\frac{d\omega}{dt} = -\frac{3k_2 GM_\odot^2 R^5}{2Ca^6 Q} \quad [56]$$

where $Q \approx 1/(\delta)$ is the dissipation or quality factor of the planet and δ the geometric lag angle between the tidal bulge and the line planet–Sun (the factor 2 stems from the fact that the main tides are semidiurnal with frequency 2ω for rapid rotators). The despin timescale τ from fast initial rotation ($\omega \gg n$) to a rotation period commensurate with the orbital period ($\omega \sim n$) can be approximated by

$$\tau = \frac{2Ca^6 Q \omega_{\text{initial}}}{3k_2 GM_\odot^2 R^5} \quad [57]$$

As a result of the strong distance dependence (a^6), Mercury and Venus can despin within the age of the solar system, whereas Mars cannot (Goldreich and Soter, 1966; see also Sections 10.04.3, 10.04.4, and 10.04.5). However, the quality factor Q of the planets is not well known, nor is its frequency (or rotation rate) dependence well understood (see, e.g., Munk and MacDonald, 1960; Lambeck, 1980; Anderson, 1989). Often, it is assumed that Q is constant for large frequencies (for fast rotation, the main tidal frequency $\sigma \approx 2\omega$) as in seismology (e.g., Dahlen and Tromp, 1998) and inversely proportional to frequency for small frequencies (the main tidal frequency can be small for $\omega \sim n$). The latter case corresponds to a constant phase lag Δt , and dissipation then vanishes when the tidal frequency goes to zero. Zero tidal frequency implies that the tidal bulge is static with respect to the planet and $\delta = 0$. During the rotational history of a planet, the torque and the despinning rate can change through the frequency-dependence of Q and as a result of internal planetary changes that can affect C and k_2 .

For slowly rotating planets, the sign of the effect of the solar tidal torque on the rotation rate depends on the obliquity, with two regions separated by an equilibrium point for the rotation rate given by

$$\omega_e = \left(\frac{2 \cos \epsilon}{1 + \cos^2 \epsilon} \right) n + \mathcal{O}(e^2) \quad [58]$$

when dissipation is assumed to be linear with frequency (or the time lag constant) (Correia and Laskar, 2003a). When $\omega > \omega_e$, the rotation rate of the planet decreases, whereas it increases for $\omega < \omega_e$. For an initially rapidly rotating planet, the rotation rate tends to decelerate to this equilibrium rotation rate.

Besides changing the rotation rate of a planet, the tidal torque also changes the obliquity as the tidal bulge is carried outside the orbital plane by rotation during the time lag for obliquities different from 0° and 180° . When the time lag is assumed to be constant and the eccentricity is negligibly small, $\epsilon = 0^\circ$ and $\epsilon = 180^\circ$ are unstable equilibrium points and $\epsilon = \arccos(2n/\omega) = \epsilon_0$ is a stable equilibrium for fast rotation ($\omega > 2n$) (Goldreich and Peale, 1970; Néron de Surgy and Laskar, 1997). Obliquity increases when $\epsilon < \epsilon_0$ and decreases otherwise. The stable equilibrium is between 0° and 90° for prograde rotation and tends to zero obliquity for a decreasing rotation rate. Tidal torques can therefore not overturn an initially prograde planet to a retrograde rotation. For slow rotation ($0 < \omega < 2n$) $\epsilon = 0^\circ$ is a stable equilibrium point and $\epsilon = 180^\circ$ an unstable point. Because tidal friction tends to drive the rotation rate ω to the slow rotation regime (eqn [58]), obliquity is eventually driven to 0° , and the rotation rate tends to the mean motion (1:1 resonance). For a noncircular orbit, the final tidally evolved state is commensurate but not synchronous with the mean orbital motion. This will be discussed further in Section 10.04.5 on Mercury.

10.04.2.4.2 Thermal atmospheric tides

A conceptually related dissipation effect on planetary rotation is due to thermal tides in the atmosphere, first shown by Lord Kelvin (Thompson, 1882). The Sun heats the atmosphere and the surface most at the subsolar point and causes pressure changes and mass motion, or thermal tides, in the atmosphere from the subsolar point to the less-heated zones. The main thermal tides of the Earth are semidiurnal and the atmospheric ‘bulge’ is almost perpendicular to the Earth–Sun direction (Chapman and Lindzen, 1970). When one takes into account a not too large response time, it follows that the solar torque on the bulge, for rapidly rotating planets ($\omega > 2n$), has an effect on the rotation rate of the planet opposite to that of the solar torque on the gravitational tidal bulge: it accelerates the rotation rate (e.g., Munk and MacDonald, 1960; Gold and Soter, 1969; Dobrovolskis and Ingersoll, 1980). Because the torque acts on the atmospheric bulge, it is assumed here that the resulting change in angular momentum of the atmosphere is transmitted to the solid body to maintain a rotation of the atmosphere not too dissimilar from that of the solid body.

The pressure perturbation in the atmosphere is inversely proportional to the tidal frequency (Chapman and Lindzen, 1970). Therefore, the effect

of thermal tides on the rotation can be neglected for fast rotation ($\omega \gg n$), and a deceleration from an initial fast rotation to a rotation commensurate with the orbital rotation is dominated by the gravitational solid body tides. Evidently, torques on the atmospheric thermal tides are mainly important for planets with a dense atmosphere, and dissipation models have mainly been developed for the Earth and Venus (e.g., Chapman and Lindzen, 1970; Dobrovolskis and Ingersoll, 1980).

For $\omega > 2n$, the thermal tides tend to bring the obliquity to either one of two stable equilibrium points, at $\epsilon = 0^\circ$ and at $\epsilon = 180^\circ$ for zero eccentricity, depending on the obliquity being smaller or larger than an unstable equilibrium obliquity value. This critical obliquity depends on ω and goes to zero for $\omega \gg n$ and to 180° for $\omega \rightarrow 2n$ (Correia and Laskar, 2003a). For large-enough initial obliquities, thermal tides can therefore tend to increase the obliquity and can overturn a planet that initially rotates rapidly in the prograde sense (Lago and Cazenave, 1979; Dobrovolskis, 1980; Yoder, 1995).

10.04.2.4.3 Core–mantle friction

Friction torques associated with differential rotation between a liquid core and a solid mantle can also change the rotation of a planet on long timescales. In our analysis of periodic rotation variations, the friction torques were neglected in a lowest-order approximation as their effects are expected to be small on short timescales. Nevertheless, friction torques have an observable effect on the nutations of the Earth (e.g., Buffett, 1992; Buffett *et al.*, 2002; Mathews *et al.*, 2002; Mathews and Guo, 2005; Deleplace and Cardin, 2006). For secular changes in the rotation, even small dissipation effects can eventually have important effects. Two different friction torques must be considered: the viscous torque due to viscosity of the core liquid, and the electromagnetic torque between core and mantle resulting from the electric currents in the conducting bottom mantle layer (see also Section 10.04.5.3). Both torques are linearly proportional to the difference of the rotation vectors of mantle and core (e.g., Rochester, 1976). For the electromagnetic torque, this can be understood by noting that the magnetic field involved in the coupling is induced by the relative velocity between the mantle and core. Therefore, usually a single friction torque is considered with parameters that include the effects of both the viscous and the electromagnetic coupling.

In response to precession, an additional flow of uniform vorticity is set up in the core (Hough, 1895; Poincaré, 1910). Without core viscosity, the vorticity of this flow is in the direction of $\vec{\omega} \times (\vec{\Omega}_P \times \vec{\omega})$, where $\vec{\Omega}_P$ is the angular velocity of precession oriented normal to the orbital plane (e.g., Stewartson and Roberts, 1963). The vorticity is thus in the $(\vec{\omega}, \vec{\Omega}_P)$ plane, and the core rotation vector is co-planar with $\vec{\omega}$ and $\vec{\Omega}_P$. Since the additional flow is steady in a precessing frame (the core co-precesses with the mantle), the change in the additional core angular momentum is in the direction of $\vec{\Omega}_P \times [\vec{\omega} \times (\vec{\Omega}_P \times \vec{\omega})]$ (Pais *et al.*, 1999). By using conservation of angular momentum, it follows that the change in the angular momentum $\vec{L}\omega \approx \vec{C}\omega$ along the rotation axis of the mantle is in the opposite direction. Because that vector is perpendicular to $\vec{\omega}$ and $\vec{\Omega}_P$, there is no corresponding change in $H = C\omega$ nor in $X = C\omega \cos \epsilon$, so neither the obliquity nor the rotation rate changes.

Core viscosity tends to decrease the relative velocity between mantle and core and to make the core precession to lag behind the mantle precession. The additional flow vorticity thus has a component in the direction perpendicular to $(\vec{\omega}, \vec{\Omega}_P)$ (Stewartson and Roberts, 1963; Roberts and Stewartson, 1965; Busse, 1968; Rochester, 1976), and the change in angular momentum $\vec{C}\omega$ has an additional component in the direction of $\vec{\Omega}_P \times (\vec{\omega} \times \vec{\Omega}_P)$, which is perpendicular to $\vec{\Omega}_P$ but not to $\vec{\omega}$. As a result, H changes in time and $X = H \cos \epsilon$ remains constant. When the rotation rate changes because of core–mantle friction, the obliquity ϵ must also change to keep the component of the spin angular momentum normal to the orbital plane constant:

$$\frac{d\epsilon}{dt} = \cot \epsilon \frac{1}{\omega} \frac{d\omega}{dt} \quad [59]$$

Since the dissipated energy must come from the rotation, core–mantle friction torques cause the rotation to slow down and the obliquity to evolve to either 0° or 180° , depending on the initial obliquity being smaller or larger than 90° (Goldreich and Peale, 1970; Rochester, 1976; Néron de Surgy and Laskar, 1997; Pais *et al.*, 1999; Correia and Laskar, 2003a; Yoder, 1995).

A main problem of core–mantle friction models is that the viscosity of the cores of terrestrial planets is poorly known (Anderson, 1989). Moreover, a turbulent boundary layer close to the CMB can develop (e.g., Yoder, 1995). Core–mantle friction becomes more important for increasing core differential

rotation with respect to the mantle during precession, and at the same time turbulence is more likely then. If the FCN period is much smaller than the precession period (such as for rapidly rotating planets as the Earth and Mars at present), the core efficiently tracks the precession of the mantle, and the effect of core–mantle friction on the long-term changes of the rotation of the planet is weak (for the effect on the Earth, see, e.g., Néron de Surgy and Laskar, 1997; Pais *et al.*, 1999). For slowly rotating planets

$$\frac{P_{\text{FCN}}}{P_{\text{Precession}}} \propto \left(\frac{n}{\omega}\right)^2 \quad [60]$$

(see eqns [4] and [47]), and the differential rotation between the core and the mantle can become larger. Therefore, core–mantle friction can become the dominant cause of secular rotation variations if rotation is small enough.

10.04.2.4.4 Climate friction

It is well known that obliquity variations induce climate changes. However, the climate can also change the obliquity of a planet by altering the mass distribution in the planet (Rubincam, 1990). In particular, during glacial cycles mass is transported between the fluid layers (atmosphere and oceans) and the polar caps. This changing surface load also induces a partially compensating internal mass redistribution. Together they lead to polar motion and small variations in the rotation rate on the timescale of the glacial cycles (Munk and MacDonald, 1960). Climate cycles also act to modify the polar flattening of the planet, and thereby the precession rate will vary. For the Earth, the oblateness can change by as much as 1% (Thomson, 1990).

Due to the fact that the mantle is viscous, the changes in the equatorial bulge will lag behind the obliquity changes and will result in a long-term effect on the obliquity. This mechanism was first described by Rubincam (1990) and called ‘climate friction’. Despite many unknowns in the climate and viscous response to obliquity oscillations, most recent analyses suggest that the climate friction effect on the long-term obliquity variations of the Earth and Mars is small compared to the planetary effect on obliquity, with obliquity changes below 1° during the age of the solar system for Mars and Earth obliquity changes smaller than 0.01° per My during the major recent Pliocene–Pleistocene glaciations (Spada and Alfonsi, 1998; Levrard and Laskar, 2003; see also Williams *et al.*, 1998).

10.04.3 Mars

10.04.3.1 Long-Term Spin Evolution

Christiaan Huygens was the first to observe that the rotation period of Mars is similar to that of the Earth. On 1 December 1659, he noted (in Latin) in his journal that “the rotation period of Mars, like that of the Earth, seems to be in a period of twenty-four hours” (see Kieffer *et al.*, 1992). The period was confirmed in 1666 by Giovanni Domenico Cassini, who determined a value of 24 h 40 min. Frederick William Herschel (1774) estimated the obliquity of Mars to be about 30°. The rotation period determined from Doppler and range measurements to the Mars Pathfinder lander and the Viking landers is 88642.66302 ± 0.00002 s, or 24h37min22.66302s ± 0.00002 s, and the obliquity $\epsilon = 25.189417 \pm 0.000033^\circ$ (Folkner *et al.*, 1997a). The inclusion of all the X-band Doppler and range-tracking data from the Mars Global Surveyor (MGS) and Odyssey missions when in orbit about Mars, except for the aerobraking part of the mission, have led to a rotation period of 24h37 min22.662997s ± 0.000007 and an obliquity $\epsilon = 25.189398 \pm 0.000011^\circ$ (Konopliv *et al.*, 2006). The spin period here is relative to the node of the Mars mean orbit and Mars mean equator of date.

Because Mars is far from the Sun and has no large moons, tidal braking has only had a small influence on its rotation (Goldreich and Soter, 1966). As the effects of thermal tides in the tenuous Martian atmosphere and of core–mantle friction for a rapidly rotating planet are even smaller (see Section 10.04.2.4), the rotation speed of Mars can be considered to be primordial (Laskar *et al.*, 2004). The dissipative effects on obliquity variations are equally small, and planetary perturbations on the orbital motion of Mars dominate the changes in obliquity. The obliquity variations depend sensitively on the precession rate (see Section 10.04.2.3), which is well known with a precision below 1%:

$$\dot{\psi} = -7582 \pm 17 \times 10^{-3} \text{ arcsec yr}^{-1} \quad [61]$$

(Konopliv *et al.*, 2006). This estimate has been determined mainly from the Doppler and range data to the Mars Pathfinder lander and the Viking landers, but also Odyssey radio-tracking data were used in the analysis. Because the current precession frequency of Mars is close to some major secular frequencies of its orbit, Mars’ obliquity performs large chaotic variations (Ward, 1974; Touma and Wisdom, 1993; Laskar and Robutel, 1993). Therefore, the evolution of

obliquity cannot be determined over longer time-scales, and the present obliquity cannot be considered as being primordial. In a statistical study of the orbital and rotational evolution of Mars, Laskar *et al.* (2004) demonstrated that the current accurately determined precession rate allows constructing reliable obliquity evolutions over 10–20 My. Sufficiently accurate initial conditions such as the precession rate are needed because the motion of Mars' spin axis is chaotic. These authors also calculated the mean and most probable obliquity values over 4 Gy: Mars seems to have a much lower obliquity now than its mean value of about 37.62° and its most probable value of about 41.80° . Obliquity and precession are important constraints on the solar insolation and hence on the climate history and the evolution of the polar caps of Mars. In particular, the present 'low' obliquity suggests that the present climate is not representative. We refer to Laskar *et al.* (2004) and references therein for an extensive study of the astronomical forcing of the Martian paleoclimate.

By using eqn [4], the polar moment of inertia C of Mars can be derived from the precession rate as $C = 0.3654 \pm 0.0008$ (Konopliv *et al.*, 2006). This measure of the radial mass concentration is one of the main and best-determined constraints on the interior structure of Mars (see Chapter 10.02).

10.04.3.2 Nutation

The observation of nutations of Mars can provide insight into the interior of Mars. If Mars behaved rigidly, nutations would only give a constraint on the moment of inertia, additionally to the constraint from precession. However, Mars evidently responds nonrigidly to forcing, and the nonrigid behavior contains important geophysical information, in particular on the core. Because the nonrigid contributions to the nutation amplitudes are currently below the limit of detectability, theoretical nutation models are used in data analyses of the rotation of Mars (Yoder and Standish, 1997; Konopliv *et al.*, 2006). In the near future, improved radio science experiments will hopefully allow distinguishing the nonrigid contributions and so offer a direct way to study properties of the deep interior of Mars. A small network of fixed landers with dual-band coherent radio transponders would be particularly suited. This mission concept has often been proposed, such as with the MESUR and INTERMARSNET projects, and the recent NetLander project (Harri *et al.*, 2000; Barriot *et al.*, 2001), but have unfortunately all

been cancelled. In this section, we review the current theoretical knowledge on Mars' nutation and discuss the geophysical potential of nutation observations.

10.04.3.2.1 Rigid nutations

The celestial mechanical problem of determining the tidal potential is most conveniently separated from the geophysical problem of calculating the rotational response to forcing by first deriving the nutations under the assumption that Mars responds as a rigid body to the gravitational (tidal) torques. Rigid nutation series for the angular momentum axis can be derived from eqn [1]. We write the equatorial components of the unit vector in the direction of \bar{H} as $(\psi + \Delta\psi) \sin \epsilon$ for the first component in the direction of the ascending node of Mars' orbital plane on Mars' equator of date and as $\epsilon + \Delta\epsilon$ for the second component in the direction of increasing obliquity. Here, $\Delta\psi$ and $\Delta\epsilon$ are the periodic nutations in longitude and obliquity. Since the equatorial components of the angular momentum are very small compared to the polar component, we approximate the magnitude of the angular momentum vector by $C\Omega$, and obtain from eqn [1]

$$\frac{d\Delta\psi}{dt} = \frac{1}{C\Omega \sin \epsilon} L_1 \quad [62]$$

$$\frac{d\Delta\epsilon}{dt} = \frac{1}{C\Omega} L_2 \quad [63]$$

where (L_1, L_2) are the equatorial components of the torque (Moritz and Mueller, 1987; Roosbeek, 1999). Equivalent expressions in a Hamiltonian theory are given in Bouquillon and Souchay (1999). For each interacting body of the solar system sufficiently far from Mars such that the torque can be approximated by its degree-two terms, we have

$$\frac{d\Delta\psi}{dt} = \frac{1}{\Omega \sin \epsilon} \frac{1}{r^3} \frac{GM(C - \bar{A})}{C} \frac{3}{2} \sin 2\delta \sin \alpha \quad [64]$$

$$\frac{d\Delta\epsilon}{dt} = -\frac{1}{\Omega r^3} \frac{GM(C - \bar{A})}{C} \frac{3}{2} \sin 2\delta \cos \alpha \quad [65]$$

where M is the mass of and r the distance to the perturbing body, and δ and α the declination and right ascension of the perturbing body, respectively. Triaxial effects of Mars are neglected here and $\bar{A} = (A + B)/2$. The equatorial torque components and the nutations can then be calculated from the ephemerides of the planets and moons (e.g., Reasenberg and King, 1979; Borderies, 1980; Hilton, 1991). The most recent and accurate nutation series (Bouquillon and Souchay, 1999; Roosbeek, 1999) are

accurate up to 0.1 mas and are based on the VSOP87 ephemerides for the planets (Bretagnon and Francou, 1988), which are valid for a few 1000 years around J2000.0 and include orbit perturbations due to the other planets. For Phobos and Deimos, the ESAPHO and ESADE ephemerides were used (Chapront-Touzé, 1990).

At a given tidal frequency, nutation can be expressed either as a sum of a sine and a cosine term in both longitude and obliquity or as a sum of two circular nutations, one prograde, the other retrograde (as compared to the rotation of the planet):

$$\begin{aligned}\Delta\epsilon + i \sin \epsilon_0 \Delta\psi &= \sum_j \left[A_j e^{i(\sigma'_j t + f_j)} + A_{-j} e^{-i(\sigma'_j t + f_j)} \right] \\ &= \sum_j \left[(A_j - A_{-j}) \cos(\sigma'_j t + f_j) \right. \\ &\quad \left. + i(A_j + A_{-j}) \sin(\sigma'_j t + f_j) \right]\end{aligned}\quad [66]$$

where ϵ_0 is the obliquity at J2000.0 (see, e.g., Demagaine *et al.*, 1995).

The largest nutation of Mars considered as a rigid body is the prograde semi-annual nutation with an amplitude of about 475 mas (Bouquillon and Souchay, 1999; Roosbeek, 1999). Semi-annual here refers to a period of half a Martian year. For the ideal case of a single biaxial planet in circular orbit about the Sun, the semi-annual nutation would be the only nutation. In that case, the torque on the planet is equal and maximal at summer and winter solstices. Other nutations are due to deviations with respect to this ideal case.

The amplitude of the semi-annual nutation is almost equal to that of the Earth. Three factors in the nutation amplitude contribute to this near equality. First, the torque depends on the third power of the distance of the planet to the Sun, which is about 1.5 times larger for Mars than for the Earth. Secondly, the nutation amplitude is proportional to the period (see eqn [18]), which is about 1.88 times longer for Mars than for the Earth. These two factors can be combined by Kepler's third law into a nutation amplitude inversely proportional to the period. Thirdly, the dynamical flattening $H_D = (C - \bar{A})/C$ of Mars is about 5/3 of the flattening of the Earth.

The amplitudes of the four main retrograde nutations are about 125, 40, 10, and 2 mas for periods of 1, 1/2, 1/3, and 1/4 year, respectively (Roosbeek, 1999; Dehant *et al.*, 2000a, 2000b). The prograde nutations at 1, 1/3, and 1/4 year have amplitudes of 91, 103, 18 mas, respectively. All these nutations are due to the gravitational torque of the Sun. The

nutations due to Mars' moons have amplitudes of about 10 mas for Phobos and 5 mas for Deimos and periods of 2.26 and 54.754 yr, respectively (Bouquillon and Souchay, 1999; Roosbeek, 1999). The largest nutations caused by the direct gravitational torque from planets are on the order of 0.1 mas and are due to Jupiter and the Earth.

Differences between the nutations for the angular momentum axis, the rotation axis, and the figure axis are at most on the order of 0.1 mas and can be neglected (Bouquillon and Souchay, 1999; Roosbeek, 1999). The largest uncertainty on the rigid nutation amplitudes is due to the observational error on the precession rate. Both precession and nutation depend linearly on the dynamical flattening H_D (Roosbeek, 1999). Any relative error on the precession rate translates into the same relative error on H_D and on the nutations. The most recent rigid nutation series both used $H_D = 0.00536$, which was derived from the very accurately known J_2 gravitational coefficient and the observationally determined precession rate of Folkner *et al.* (1997a). The 0.5% error of Folkner *et al.* (1997a) results in an error of 2.4 mas for the main nutation. With the new and more accurate precession rate of Konopliv *et al.* (2006), the rigid nutation amplitudes would increase by 0.3%, and the uncertainty on the main nutation reduces to 1.6 mas.

10.04.3.2.2 Nonrigid nutations

A nonrigid nutation series is derived from a rigid nutation series by means of a transfer function $T(\sigma)$ which describes the ratio of the nonrigid nutation amplitude to the rigid nutation amplitude for frequency σ (Dehant *et al.*, 2000a). We will use a transfer function for a planet with a liquid core because Mars most likely has, at least partially, a liquid core. From Mars Global Surveyor tracking, Yoder *et al.* (2003) derived a value of 0.153 ± 0.017 for the Love number k_2 , which is clearly larger than the value of about 0.07 for an entirely solid Mars (e.g., Van Hoolst *et al.*, 2003; Yoder and Standish, 1997). The liquidity of the core is confirmed by Konopliv *et al.* (2006), who obtained a best estimate of $k_2 = 0.152 \pm 0.009$. Thermal evolution models also point to an at least partially liquid core (Stevenson *et al.*, 1983; Schubert and Spohn, 1990). For a planet with a solid mantle and a liquid core, the transfer function can be derived from Liouville equations [22]–[24] and [31]–[33] as

$$T(\sigma) = \frac{(\sigma - \sigma_E)(A\sigma' - A^m\sigma'_{FCN})}{A^m(\sigma - \sigma_{CW})(\sigma - \sigma_{FCN})}\quad [67]$$

which shows that nutations are resonant at frequencies equal to the CW frequency and the FCN frequency. For a planet containing also a solid inner core, resonances with the FICN and the ICW must additionally be considered. The FCN is particularly important because it has a long retrograde period in an inertial reference frame between about 200 and 300 days, depending on the interior structure, close to the periods of the main nutations (see eqn [47], Hilton, 1992; Yoder and Standish, 1997; Van Hoolst *et al.*, 2000b). Therefore, the FCN can have a large effect on nutation. For the Earth, relatively large resonant nutation amplifications have been detected and have been used to demonstrate that the flattening of the core is larger than that of a hydrostatic Earth based on the PREM model (Gwinn *et al.*, 1986; Mathews *et al.*, 1991, 2002).

Dehant *et al.* (2000a) calculated the Mars nutation transfer function by numerically solving the system of differential equations describing the reaction of Mars to external gravitational forcing. For a detailed review of this approach developed for the Earth, we refer to Chapter 3.10. In Figure 2, the relative enhancement of the nutation amplitudes with respect to the rigid amplitudes is shown for models with a liquid core (see Dehant *et al.*, 2000a). The amplitude amplification due

to the resonance with the FCN is striking. For interior structure model A of Sohl and Spohn (1997), the amplifications are 1.4%, 6%, 30%, 7%, and 1.1% for the retrograde nutations at 1, 1/2, 1/3, 1/4 years, and the semiannual prograde nutation, respectively. The amplifications, and amplitudes, of the retrograde nutations depend sensitively on the model. For example, for a model with a 200 km larger core than Sohl and Spohn's model A, the amplifications are 2.0%, 8%, 500%, 15%, and 1.7% (Dehant *et al.*, 2000b).

The amplification of the largest nutation with prograde semi-annual period depends much less on the FCN period since it is much further from the retrograde FCN period, and thus from the resonance peak. The amplification with respect to the rigid nutation term is about 1% or 5 mas, for any realistic model of Mars' interior. Future determinations of the amplitudes of nutation terms with a radio science experiment are expected to have a precision on the order of a mas (Dehant *et al.*, 2000a). Only for the prograde semi-annual nutation is the amplification always above the observational precision and larger than the uncertainty on the rigid amplitude, irrespective of the interior structure of Mars. Therefore, this nutation is ideally suited for further independent evidence of the liquidity of the core.

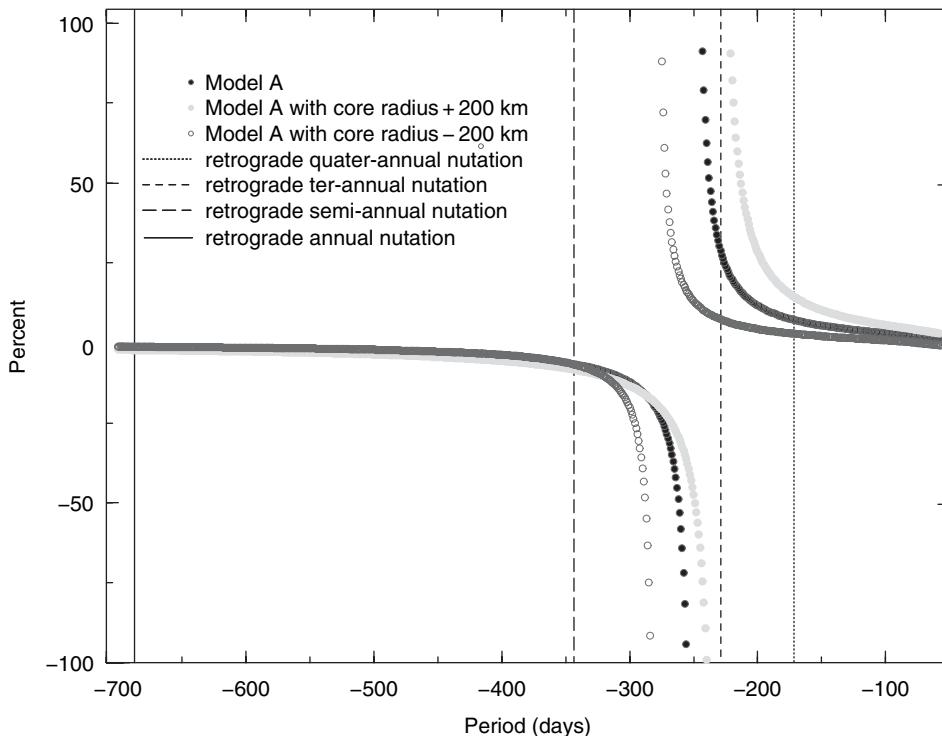


Figure 2 Relative differences between nonrigid and rigid nutation amplitudes as a function of period (retrograde band).

10.04.3.2.3 Nutation and interior structure

Besides establishing the liquidity of the Martian core, the nutations can be used to constrain other properties of the interior of Mars, by making use of the resonances of the main nutations with rotational normal modes whose frequencies depend sensitively on Mars' interior. The main resonance is with the FCN (see Figure 2), but also the FICN could resonantly enhance the nutations. The resonant amplification of the nutations depends strongly on the normal-mode periods. Therefore, observation of the amplifications allows the determination of the normal-mode frequencies.

The FCN period depends mainly on the CMB polar flattening α_f because the restoring force for this normal mode, the inertial pressure force on the CMB, is proportional to that flattening. Therefore, the FCN period decreases with increasing CMB flattening. For models with increasing core radius from 1268 to 1768 km and the same mantle profile as model A of Sohl and Spohn (1997), the FCN period decreases almost linearly from 287 to 230 days, or about 11.5 days per 100 km (see Figure 3). This is almost fully due to the increase in CMB flattening, α_f , as a function of core radius, itself a consequence of the larger core density changes in this sequence of models (Van Hoolst *et al.*, 2000a). The FCN period therefore seems to be a good indicator of core density. Because of the relatively small mass of the core ($\sim 15\%$ of the total mass), it is then also a good indicator of the density jump at the CMB. An almost linear increase of the period of about 22 days in absolute values was found for a 1 g cm^{-3} increase of the density jump at the CMB (Van Hoolst *et al.*, 2000a).

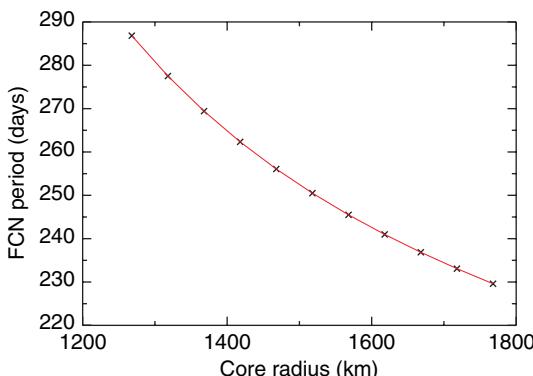


Figure 3 FCN period as a function of core radius.

Because the FCN eigenfunction represents essentially rotational motion, the influence of the velocities V_p and V_s and of mantle inelasticity on the FCN period is small. Hilton (1992) considered variations in rigidity of 5% and in incompressibility of 2% and found relative differences of the order of 10^{-5} , or less than a minute for the FCN period. Van Hoolst *et al.* (2000a) considered larger changes of 10% and 21% in rigidity and incompressibility and showed that the FCN period decreases for increasing velocities by about 1 day and 2 days for changes of the velocities of 5% and 10%, respectively. The inelasticity of the mantle, even for small and probably unrealistic Q -values down to 50, increases the FCN period by less than 1 day (Van Hoolst *et al.*, 2000b). The effect of triaxiality of Mars on the FCN period is also very small. Inclusion of the equatorial flattening in the theoretical calculations lengthens the FCN period by less than 0.5 days (Van Hoolst and Dehant, 2002). Given these small effects for hydrostatic models of Mars, we can conclude that the FCN period is a good proxy for core density or, even better, for the CMB density jump.

The closer the FCN period is to one of Mars' main nutations, the more accurately the FCN period and thus the core radius and density of Mars can be determined. After 50 weeks of Doppler tracking observations, with a precision of 1 mm s^{-1} , between four landers and an orbiter, and the orbiter and the Earth, the FCN period is expected to be determined with a 10 day uncertainty (Yseboodt *et al.*, 2003).

Mantle convection and the presence of a solid inner core could complicate the interpretation of the nutation amplitudes. Mantle convection can deform the CMB and change the core flattening. The Tharsis region is the dominant topography on Mars and could be associated with a hot upwelling plume in the mantle (Harder and Christensen, 1996; Kiefer *et al.*, 1996). Temperature- and pressure-dependent convection models indicate that a mantle plume could contribute at most about 10% to the geoid, suggesting that the Tharsis rise is supported by flexure (Roberts and Zhong, 2004). Recent lava flows could be due to upwelling (Kiefer, 2003), but could also be related to partial melting under a thick crust with a lower thermal conductivity than that of the mantle (Schumacher and Breuer, 2005). As Tharsis is situated in an equatorial area, it increases the polar flattening of the core. Steady-state convection models that account for the geoid high at Tharsis show that the FCN frequency can be decreased down to

about 100 days, in particular, when no lower mantle perovskite layer is present (Defraigne *et al.*, 2001). The FCN resonance then shifts to smaller rigid nutation terms (see **Figure 2**).

Yoder *et al.* (2003) showed that at least the outer part of the core is liquid, but it is not known whether there would be a solid inner core, nor how large it would be. The main uncertainties are the temperature profile and the concentration of light elements in the core, especially sulfur, which has a large effect on the melting temperature of the core assemblage (Schubert and Spohn, 1990; Spohn, 1991). Using SNC meteorites, Dreibus and Wänke (1985) showed that the percentage of sulfur in the core is about 14 wt.%. On the basis of another analysis of SNC meteorites, Gaetani and Grove (1997) obtained sulfur concentrations as low as 0.4 wt.%. For a given pressure, the lower the sulfur content is, the higher the temperature at which the core solidifies. As a consequence, the time when the core starts to solidify depends strongly on its composition. A high sulfur content tends to keep the core entirely liquid (Schubert and Spohn, 1990; Schubert *et al.*, 1992; Breuer and Spohn, 2003).

The absence of a magnetic field of global scale implies that dynamo action in the Martian core is presently not active. As inner core growth is very efficient in driving a dynamo, the magnetic field observations have been used to argue against the existence of a solid inner core (e.g., Breuer and Spohn, 2003). However, core growth could be too small to sustain a dynamo, or could have stopped due to core reheating after an early period of global plate tectonics (Stevenson, 2001). Besides dynamo studies and thermal evolution calculations, observations are needed to decide upon the existence of an inner core. Here, we show that nutations could provide constraints on a solid inner core.

At the Martian core pressures and for a sulfur concentration below the eutectic concentration, mainly iron solidifies when the temperature decreases below the liquidus temperature of the core assemblage and forms a quasi pure iron solid inner core. The sulfur concentration of the outer core therefore gradually increases with time, until the outer core reaches the eutectic composition. From then on, the freshly solidified material has the same eutectic composition as the liquid. When the outer core has a sulfur concentration below the eutectic value, the FCN period almost does not change, and the inner core influence on the nutations is below 1 mas (Van Hoolst *et al.*, 2000b; Defraigne *et al.*, 2003;

Dehant *et al.*, 2003). The inner core and its moment of inertia are too small to noticeably influence the FCN, which is essentially a differential rotation of the outer core. The effect of the FICN on nutations is then also negligible, except if the FICN period happens to be very close to the prograde annual or semi-annual nutation (Dehant *et al.*, 2003; Defraigne *et al.*, 2003). Observation of such an amplification would be a strong indication for the existence of an inner core, but would require a lucky coincidence. For progressively larger inner core, the FCN period decreases and the resonance shifts to higher retrograde nutations, which have much smaller rigid amplitudes. Except for very large amplifications, these nutations are below 1 mas. The amplification of the largest nutation, the prograde semi-annual, also increases for larger inner cores. However, this amplification due to the FCN resonance is almost canceled by the resonance with the FICN, whose period decreases rapidly to zero in the limit of an entirely solid core (Defraigne *et al.*, 2003). Failure to detect amplification of the semi-annual nutation with a future sufficiently accurate geodesy experiment could then be interpreted as evidence for a large inner core.

10.04.3.3 Rotation Speed Variations

The atmosphere of Mars consists of 95% carbon dioxide (CO_2) (Owen *et al.*, 1977), which condenses during the winter seasons at high latitudes to form periodic polar CO_2 caps. During spring and summer seasons, the caps sublime and the mass is again redistributed in the atmosphere (James *et al.*, 1992). In this seasonal cycle of mass transfer between the atmosphere and the polar caps, a large fraction of the total mass of the atmosphere condenses at the polar regions. The maximum mass of the north and south seasonal caps is $3.0 \pm 1.0 \times 10^{15}$ and $5.5 \pm 1.0 \times 10^{15}$ kg, respectively (Karatekin *et al.*, 2006b; Konopliv *et al.*, 2006; Litvak *et al.*, 2006), or about one-quarter of the total mass of the atmosphere of about 2.5×10^{16} kg (Lodders and Fegley, 1998). Because of conservation of angular momentum, this large mass redistribution by seasonal sublimation and condensation of CO_2 and the (associated) winds can induce seasonal changes in the rotation of Mars. The more the mass is concentrated in the polar regions, as solid CO_2 ice in the polar caps, the smaller the moment of inertia C about the polar axis is and the faster the Mars rotates. When the polar caps sublime, the CO_2 is distributed more

uniformly over the Martian surface and Mars rotates slower. Qualitatively, the effect on rotation scales as the ratio of the net redistributed mass to the planet's total mass (Chao and Rubincam, 1990). This ratio is about 1.3×10^{-8} , at least one to two orders of magnitude larger than any mass redistribution on Earth, suggesting large rotation rate variations. On the other hand, the redistribution of matter is mainly in the north/south direction. If the growth of one polar cap would be symmetrical with respect to the retreat of the other polar cap, there would be no net change in the moment of inertia of Mars, and only winds in the atmosphere could change the rotation rate. In reality, the caps do not grow symmetrically. Since the orbit of Mars is quite eccentric (eccentricity equal to 0.0934), the temperatures at the two poles differ for a given season, and the seasons do not have the same length on both hemispheres. For example, winter on the southern hemisphere lasts almost a month longer than on the northern hemisphere (Kieffer *et al.*, 1992). Differences in polar topography, thermal inertia, and albedo further contribute to asymmetry between the increase and decrease of the two polar caps.

Seasonal changes in the rotation speed of Mars have first been detected from radio tracking data to the Viking landers (Reasenberg *et al.*, 1980; Folkner *et al.*, 1997a; Yoder and Standish, 1997). Although the rotation rate variations can best be determined from a lander, more recent solutions are based on Mars Global Surveyor (MGS) and Odyssey tracking data, which are both qualitatively and quantitatively better than the older Viking data. Solutions based on the orbiter tracking data are consistent with the lander data within the uncertainties (Konopliv *et al.*, 2006). Using both lander and orbiter data, Konopliv *et al.* (2006) determined amplitudes of 456 ± 51 , 169 ± 43 , 31 ± 35 , and 17 ± 31 mas for the rotation angle variations at 1, 1/2, 1/3, and 1/4 year period, respectively. In these studies, a theoretical nutation model is included. If it were omitted, it would substantially change the estimates of the seasonal rotation variations (see, e.g., Yoder *et al.*, 2003).

Instead of rotation angle variations, relative changes in the rotation rate $\delta\omega_3/\Omega = m_3$, changes in the length of the day ($\text{LOD}(t) = 2\pi/\omega_3(t)$), or variations in the time needed for the planet to rotate over the angle $\phi(t)$ at constant angular velocity Ω are often used to describe rotation variations. The latter variations are denoted as δUT by analogy with LOD

variations for the Earth. We have the following useful relations between the rotation variations:

$$\frac{\Delta\text{LOD}(t)}{\text{LOD}} = -m_3(t) \quad [68]$$

$$\delta\phi(t) = \Omega \int_{t_0}^t m_3(t') dt' \quad [69]$$

$$\delta\text{UT}(t) = \int_{t_0}^t m_3(t') dt' \quad [70]$$

$$= \frac{1}{\Omega} \delta\phi(t) \quad [71]$$

$$= -\frac{\Omega}{2\pi} \int_{t_0}^t \Delta\text{LOD}(t') dt' \quad [72]$$

For a periodic variation with frequency ω , the amplitudes A_{LOD} , A_{UT} , A_{m_3} , and A_ϕ of the variations in LOD, UT, m_3 , and ϕ , respectively, are related by

$$A_{m_3} = \omega A_{\text{UT}} = \frac{\Omega}{2\pi} A_{\text{LOD}} = \frac{\omega}{\Omega} A_\phi \quad [73]$$

As an example, the amplitude of the annual rotation variations estimated by Konopliv *et al.* (2006) as $A_\phi = 456$ mas corresponds to amplitudes $A_{\text{UT}} = 31.2$ ms, $A_{m_3} = 3.30 \times 10^{-9}$, and $A_{\text{LOD}} = 0.293$ ms. For the semi-annual amplitude, we have $A_\phi = 169$ mas, $A_{\text{UT}} = 11.6$ ms, $A_{m_3} = 2.45 \times 10^{-9}$, and $A_{\text{LOD}} = 0.217$ ms. Hereafter, only rotation variations in mas will be quoted.

With modern radio tracking of a lander over at least one Martian year, the accuracies on the rotation variations could be increased. Because the older Viking data are an order of magnitude less precise than modern X-band tracking data and the rotation estimates based on them are of comparable accuracy to those based on the MGS and Odyssey data, an order of magnitude improvement can be expected. Direct two-way range and X-band Doppler measurements between an Earth tracking station and a lander on Mars, or Earth-based radio-tracking observations of an orbiter that is itself tracking one or more landers would enable determination of the rotation variations with an accuracy of a few mas (Folkner *et al.*, 1997b; Barriot *et al.*, 2001; Yseboodt *et al.*, 2003).

Theoretical estimates of rotation rate variations were first derived by Philip (1979), who used Viking data to calculate the atmospheric friction torque associated with zonal winds. This torque transfers angular momentum between the atmosphere and the solid planet and causes rotation rate variations. Although the torque approach gives insight into the mechanisms transferring angular momentum

between the solid body and the atmosphere, most studies have used an angular momentum approach as developed in Section 10.04.2 since this has proven to be more reliable (see also Sanchez *et al.*, 2003). Wind and loading data to calculate the excitation functions or the angular momentum functions of Section 10.04.2 can be obtained from general circulation models (GCMs) of the Martian atmosphere or from physical considerations on mass redistribution within the Martian atmosphere and polar caps (see, e.g., Chao and Rubincam, 1990; Yoder and Standish, 1997). Recent state-of-the-art general circulation models were used by Van den Acker *et al.* (2002) and Sanchez *et al.* (2003). The former study used the Martian Climate Database v3 from the Laboratoire de Météorologie Dynamique (LMD) in Paris, France (Forget *et al.*, 1999; Lewis *et al.*, 1999), the latter the NASA Ames GCM (Pollack *et al.*, 1990).

The estimated annual and semi-annual rotation angle amplitudes are 394 and 192 mas, respectively, in the study of Van den Acker *et al.* (2002) and agree with the observational values of Konopliv *et al.* (2006) within their uncertainties. The annual and semi-annual amplitudes of Sanchez *et al.* (2003) are 291 and 106 mas, respectively. The latter study did not include effects of elastic deformation and of a liquid core. The loading effect of Van den Acker *et al.* (2002) is in better agreement with the theoretical estimates of Yoder and Standish (1997), who used a simpler model for the atmospheric and ice cap loading (see also Chao and Rubincam, 1990) but also included effects of elastic yielding and a liquid core, suggesting that these effects could explain the difference with Sanchez *et al.* (2003). The deformational effect decreases the rotation variations by about 8% (Defraigne *et al.*, 2000). As the ratio C/C^m is only about 1.03 (the value for the standard model A of Sohl and Spohn (1997); see, e.g., Van Hoolst *et al.*, 2000b), a liquid core increases LOD variations at most by about 3%. However, this effect is to a large extent compensated by the different deformational effect (k'_2 in eqn [44]). Deformation tends to decrease rotation variations, and its effect on rotation is a few percent larger for a liquid core than for a solid core. As a result, the effect of the state of the core is very small and on the order of 1% (Defraigne *et al.*, 2000, Karatekin *et al.*, 2006a). The differences between the results of Van den Acker *et al.* (2002) and Sanchez *et al.* (2003) can therefore not be explained in terms of the neglect of these effects by Sanchez *et al.* (2003), but are most likely due to differences in the atmospheric models. This indicates that ΔLOD observations are useful to constrain the GCMs. Note that rotation

differences between GCM models having different atmospheric dust contents and dust storms with respect to the standard model for the ‘mean’ climate of Mars are about one order of magnitude larger than the expected accuracy of future ΔLOD measurements (Defraigne *et al.*, 2000; Van den Acker *et al.*, 2002).

The separate effects of wind, atmospheric loading, and ice cap loading can be studied with the angular momentum approach. The winds have the smallest effect on the rotation rate of Mars in contrast with the Earth where the winds are the dominant source for seasonal LOD variations (Cazenave and Balmino, 1981; Yoder and Standish, 1997; Defraigne *et al.*, 2000; Van den Acker *et al.*, 2002). From data of a general circulation model of the Martian atmosphere developed at the Laboratoire de Météorologie Dynamique (LMD) in Paris, France (Forget *et al.*, 1999), wind is found to induce rotation angle variations with an amplitude of 14 mas for the annual period and 76 mas for the semi-annual period (Van den Acker *et al.*, 2002). The contributions from atmosphere and ice cap loading are almost out of phase, with the atmosphere having an effect almost three times as large as that of the ice caps (Defraigne *et al.*, 2000; Van den Acker *et al.*, 2002). The effect of the ice caps is smaller than that of the atmospheric loading because mass at the polar ice caps has only a small contribution to the polar moment of inertia C , whereas atmospheric masses can be a Mars radius further from the polar axis and cause a larger change in the polar moment of inertia.

By neglecting the wind contribution and the small effects on rotation variations due to the elastic yielding of Mars and to the presence of a liquid core, LOD variations can most simply be obtained from the conservation of angular momentum $C\omega_3$. Here we assume principal axis rotation, which is correct up to the first order in the rotation variations (see Section 10.04.2). The variation in the relative rotation rate is then proportional to the relative change in the largest moment of inertia C :

$$\frac{\Delta\omega_3}{\Omega} = -\frac{\Delta C}{C} \quad [74]$$

From the definition of the gravitational coefficient $J_2 = [C - (A + B)/2]/MR^2$ and relation [68] between rotation rate variations and LOD variations, we have

$$\frac{\Delta\text{LOD}}{\text{LOD}} = \frac{2}{3} MR^2 \frac{\Delta J_2}{C} \quad [75]$$

LOD variations can thus directly be determined from estimates of $\Delta\mathcal{J}_2$. The contribution of the atmosphere to $\Delta\mathcal{J}_2$ is about an order of magnitude smaller than that of the ice caps (Chao and Rubincam, 1990; Smith *et al.*, 1999). LOD variations can then be determined from polar ice cap data only. In the approach with angular momentum functions described above, the atmosphere pressure is more important than the polar caps. This seemingly contradictory finding is due to the fact that the difference between the terms in the $\Delta\mathcal{J}_2$ approach indicates a difference in \mathcal{J}_2 , but not in the rotation variations. To relate $\Delta\mathcal{J}_2$ to rotation variations, mass conservation is assumed, which is not satisfied for the atmosphere or the ice caps separately. By using eqn [75], observational estimates of LOD variations can be obtained from several independent data. First, MGS and Odyssey radio-tracking data have been used to estimate variations in the zonal degree-two coefficient of the gravitational field $\Delta\mathcal{J}_2$ (Smith *et al.*, 2001; Yoder *et al.*, 2003; Konopliv *et al.*, 2006). Secondly, the ice cap contribution to $\Delta\mathcal{J}_2$ can be estimated from measurements of the onboard laser altimeter MOLA of MGS (Smith *et al.*, 2001) and the High Energy Neutron Detector (HEND) of Mars Odyssey (Litvak *et al.*, 2004). Cap elevation changes up to 1.5–2 m were found from MOLA data (Smith *et al.*, 2001; Aharonson *et al.*, 2004). LOD variations based on MOLA data were calculated by Karatekin *et al.* (2006a) who used a cap density of 1000 kg m^{-3} , as suggested by Smith *et al.* (2001) to approximately match the amplitude of the observed time-variable gravity signal. Karatekin *et al.* (2006a) also calculated LOD variations from estimates of the surface density of seasonal CO₂ deposits at different latitudes determined from HEND measurements of seasonal variations in neutron flux (Litvak *et al.*, 2004). The ΔLOD obtained from tracking data of the MGS spacecraft, and measurements of the seasonal CO₂ polar deposits show large differences (of several tens of percent) both in amplitudes and phases, due to large measurement errors and modelling uncertainties in all methods (Karatekin *et al.*, 2006a).

Besides loading and wind effects, also tides can induce rotation variations. Zonal tides change the moment of inertia C and so, by conservation of angular momentum, the LOD changes. From the change in C for a model with a decoupled liquid core (see also Yoder *et al.*, 1981), amplitudes of the annual and semi-annual tidally driven rotation angle variations were estimated by Yoder and Standish (1997) to be about 50 and 30 mas, respectively, or about one-tenth

of the amplitudes induced by atmosphere and ice cap loading. Defraigne *et al.* (2000) numerically integrated the equation of motion, Poisson's equation, and the continuity equation from the center of the planet up to the surface (see Defraigne and Smits, 1999) to determine the rotational response of the planet to the zonal tides accurately determined by Roosbeek (1999). Elastic as well as inelastic mantles were considered, and the annual and semi-annual rotation angle amplitudes are 15 and 8 mas, respectively. The difference between both models is on the order of 1 mas, at the limit of the expected future observational precision.

10.04.3.4 Polar Motion

10.04.3.4.1 Seasonal polar motion

Atmospheric dynamics and the condensation/sublimation cycle of the atmosphere and polar caps not only change the rotation speed, but also cause fluctuations in the orientation of the figure axis with respect to the rotation axis of Mars, partly because the polar caps are not symmetrical with respect to the rotation axis. Up to now, polar motion of Mars has not been measured directly. Model calculations have shown that the seasonal polar motion amplitudes are at most 20 mas, which is below the present-day detectability limits, but larger than the expected future precision of a few mas (e.g., Barriot *et al.*, 2001; Defraigne *et al.*, 2000; Yseboodt *et al.*, 2003).

The main atmospheric forcing is at seasonal frequencies, resulting in largest wobble amplitudes of about 10 mas at 1, 1/2, and 1/3 year, the exact values depending on the atmospheric and internal structure models used (Chao and Rubincam, 1990; Yoder and Standish, 1997; Defraigne *et al.*, 2000; Van den Acker *et al.*, 2002; Sanchez *et al.*, 2004). The large amplitude at 1/3 year is most likely due to the resonance with the Chandler wobble with a period of about 200 days (eqn [46], Yoder and Standish, 1997; Van Hoolst *et al.*, 2000b; Dehant *et al.*, 2003; Zharkov and Gudkova, 2005; Konopliv *et al.*, 2006).

Based on the angular momentum approach of Section 10.04.2 and using the state-of-the-art GCM of the LMD (eqn [45]), Van den Acker *et al.* (2002) obtained annual and semi-annual amplitudes of 12.0 and 9.8 mas, respectively. The contribution of the atmospheric pressure is about twice that of the contribution of the ice caps. The wind-induced polar motion has about the same amplitudes as the polar motion due to the ice caps. The NASA Ames GCM has been used by Sanchez *et al.* (2004) to estimate

polar motion series. These authors used the angular momentum approach but did not include a fluid core in their model, except for the deformational effects due to the incremental centrifugal acceleration. This effect is proportional to the Love number k_2 , for which Sanchez *et al.* (2004) used the value determined by Yoder *et al.* (2003), who showed that this value is a strong indication that the core is, at least partially, liquid. The annual and semi-annual polar motion is less than half that in Van den Acker *et al.* (2002), with amplitudes of 5.1 and 2.5 mas, respectively. The atmospheric contribution, including both loading and wind, is twice as large as the ice cap contribution for the annual term and about equal for the semi-annual contribution. Sanchez *et al.* (2004) also calculated higher harmonics, and found amplitudes of 4.6, 2.0, and 1.2 mas at the 1/3, 1/4, and 1/5 year periods, respectively. The relatively large value at 1/3 year, as in Yoder and Standish (1997), is probably due to the closeness to the Chandler wobble period.

If Mars has a solid inner core, also the ICW period could be close to a seasonal forcing frequency, resulting in resonant polar motion amplification (Dehant *et al.*, 2003). The degree of amplification is difficult to estimate, and depends among others on the Q -factor of the ICW, which is unknown. Nevertheless, the resonance strength for the ICW is very small, and a very close approach is needed for noticeable effects (Dehant *et al.*, 2003, 2006).

Konopliv *et al.* (2006) determined the Martian wobble from their estimates of the order 1, even-degree $C_{l,1}$ and $S_{l,1}$ gravity coefficients. In their procedure, basic models for the asymmetric shape of the ice caps are used to separate the degree-two from the higher degrees terms. As the method is based on the variation of the gravity coefficients, the wind contribution to polar motion is neglected, although that contribution could be as large as the ice cap contribution to polar motion (Van den Acker *et al.*, 2002). The annual and semi-annual wobble amplitudes, respectively, 13.6 and 11.2 mas, agree remarkably well with the theoretical values of Van den Acker *et al.* (2002).

10.04.3.4.2 Chandler wobble

Besides polar motion at seasonal frequencies, the Martian atmosphere could also excite the Chandler wobble. For the Earth, the Chandler wobble is the largest polar motion signal and is thought to be mainly excited by ocean-bottom pressure fluctuations (Gross, 2000). Because Mars has no oceans and the Chandler wobble period for Mars

($P_{\text{CW}} \approx 200$ days) is far from the annual and semi-annual periods where most excitation power of the atmosphere is expected (Sanchez *et al.*, 2004; Dehant *et al.*, 2006), the Chandler wobble is not expected to have a large amplitude. From the spectral amplitude in the pressure recordings of the Viking landers near the Chandler wobble period, Yoder and Standish (1997) estimated a Chandler wobble amplitude of about 10 mas, and possibly as large as 50 mas with ice loading excitation included. Dehant *et al.* (2006) estimated the power spectral density of the atmospheric excitation close to the CW period from the LMD GCM (Forget *et al.*, 1999), and determined a CW amplitude between 30 and 100 mas for values of the quality factor Q between 80 and 300 (for estimates of the Q factor, see Smith and Born, 1976; Yoder, 1982; Bills *et al.*, 2005). Other excitation sources such as Marsquakes are too small (Golombek *et al.*, 1992; Golombek, 2002) to excite polar motion to an observable level (Dehant *et al.*, 2006).

From the observed variations in the order 1 gravity coefficients, and using ice cap models, Konopliv *et al.* (2006) determined an upper bound on the CW amplitude of about 10 mas. In their analysis based on MGS and Odyssey radio-tracking data, the CW signal in the gravity variations could not be isolated from the 1/3 yr forced wobble. Unambiguous detection of the CW and determination of its period could be used to determine interior properties of Mars, although the CW period is mainly constrained by the global dynamical polar and equatorial flattenings of Mars. For an entirely rigid Mars, the CW period (Konopliv *et al.*, 2006) would be

$$P_{\text{CW}}^{\text{rigid}} = \frac{2\pi}{\Omega} \sqrt{\frac{AB}{(C-A)(C-B)}} = 190.8 \pm 0.3 \text{ days} \quad [76]$$

much smaller than the corresponding value for the Earth (about 304 days) because of the larger flattening of Mars. The true CW period is most likely between about 190 and 210 days (Yoder and Standish, 1997; Van Hoolst *et al.*, 2000b; Dehant *et al.*, 2003; Zharkov and Gudkova, 2005; Konopliv *et al.*, 2006). With some more years of radio-tracking data to Mars orbiters, the CW period will probably be resolved (Konopliv *et al.*, 2006). The CW only weakly depends on the core and on the inner core. The core can have an influence on the order of a few days (Van Hoolst *et al.*, 2000b; Konopliv *et al.*, 2006), and the changes induced by the inner core are smaller than the uncertainty on the global dynamical

flattening (Dehant *et al.*, 2003). Large mantle anelasticity (Q -values below 100) could increase the period by up to about 7 days (Van Hoolst *et al.*, 2000b). Uncertainties in the rigidity and compressibility profiles also affect the Love number k_2 and therefore the CW period (see eqn [46]). For large changes of 10% and 21% in the mantle seismic velocities V_p and V_s with respect to a standard model, the CW period changes by about 0.5 and 5 days, respectively (Van Hoolst *et al.*, 2000a). To derive interior properties of Mars from the CW period, its future observational accuracy must therefore be at least about 1 day.

10.04.4 Venus

10.04.4.1 Long-Term Spin Evolution

Analyses of radar observations of Venus during its inferior conjunction in fall 1962 showed that Venus rotates retrogradely with a period of about 250 days about an axis nearly perpendicular to its orbit (Carpenter, 1964; Goldstein, 1964). This period is close to the 243.16 day period for the synodic resonance with the orbital motion of the Earth in which Venus would show exactly the same face to the Earth at each inferior conjunction and would rotate exactly five times as seen from the Sun between consecutive inferior conjunctions, suggesting that Venus could indeed occupy that resonance (Goldreich and Peale, 1966; Gold and Soter, 1969). Further observations with the Goldstone, Arecibo, and Haystack radar systems yielded an improved period of 243.026 ± 0.006 days (Carpenter, 1966, 1970; Dyce *et al.*, 1967; Shapiro *et al.*, 1990), and analyses of Magellan radar images and X-band radio-tracking data have led to an averaged rotation period of 243.0200 ± 0.0006 days (Konopliv *et al.*, 1999, see also Davies *et al.*, 1992; Konopliv *et al.*, 1993), both very close to but not exactly equal to the resonance period of 243.16 days.

The long rotation period as well as the retrograde rotation and the almost perpendicular position of Venus with respect to its orbital plane could be primordial. More plausibly, however, the rotation of Venus has changed over time. Dissipation of tides in Venus raised by the Sun causes a delay in the appearance of the tidal bulge of Venus, and the solar torque on it tends to decrease the rotation period of Venus and to drive Venus to a 1:1 resonance with its orbital motion on a timescale of 10^8 to 10^9 years (Dobrovolskis and Ingersoll, 1980; Yoder, 1995; Correia and Laskar, 2003a). Venus' slow rotation

therefore cannot be taken as an indication for a primordial slow rotation. Instead of being locked in a spin-orbit or spin-synodic resonance, the rotation of Venus is thought to be the result of a balance between solid body tidal torques, which drive Venus to synchronous rotation, and atmospheric torques, which drive it away (Gold and Soter, 1969; Kundt, 1970; Lago and Cazenave, 1979; Dobrovolskis, 1980; Dobrovolskis and Ingersoll, 1980; Shen and Zhang, 1989; McCue and Dorman, 1993; Yoder, 1995; Correia and Laskar, 2001, 2003a, 2003b; Correia *et al.*, 2003). For the Earth, the thermal tides are too small to balance the tidal gravitational dissipation, but Venus has a much denser atmosphere and, most likely, a smaller tidal dissipation due to the absence of oceans. The acceleration of the Earth's rotation due to the thermal tides is about three times smaller than the deceleration due to the solar gravitational torque, and less than 10% of the total solid body gravitational effect including the Moon (Munk and MacDonald, 1960).

For a fast initial rotation ($\omega \gg n$), both the gravitational tides and friction at the core–mantle boundary decelerate the rotation of Venus, whereas the atmospheric thermal tides that tend to accelerate Venus are small and their contribution to the rotation rate can be neglected. Once the slow rotation regime ($\omega \sim n$) has been reached, the core–mantle friction torque dominates the obliquity evolution and drives Venus to either one of the two stable equilibrium points at $\epsilon = 0^\circ$ or $\epsilon = 180^\circ$, for a wide range of values for the dissipative parameters (Goldreich and Peale, 1970; Yoder, 1995; Correia *et al.*, 2003). For these final obliquities, core–mantle friction does not affect the rotation rate anymore, and thermal atmospheric tides can balance the solid body tidal braking effect. Only four final stable rotation states are possible: a retrograde rotation at obliquity $\epsilon = 0^\circ$, a retrograde rotation at obliquity $\epsilon = 180^\circ$, a prograde rotation at obliquity $\epsilon = 0^\circ$, and a prograde rotation at obliquity $\epsilon = 180^\circ$ (Correia and Laskar, 2001; Correia *et al.*, 2003). The two retrograde states can represent the observed present rotation of Venus and correspond to different rotation histories. If Venus had a prograde initial rotation, in a first scenario Venus' rotation axis flips direction under the influence of atmospheric torques and slows down (Lago and Cazenave, 1979; Dobrovolskis, 1980; Shen and Zhang, 1989; McCue and Dorman, 1993; Yoder, 1995). In the second scenario, Venus slows down while changing its obliquity to 0° , and then develops a retrograde rotation (Kundt, 1970; Correia *et al.*, 2003). The current

rotation state of Venus therefore places no constraints on the primordial spin.

Correia and Laskar (2001, 2003a) have extensively tested the long-term evolution of the spin of Venus in realistic simulations that include the planetary perturbations. Without planetary perturbations, a final rotation state with $\epsilon = 180^\circ$ is only possible for sufficiently large initial obliquities and a dominant atmospheric torque (Lago and Cazenave, 1979; Dobrovolskis, 1980; Shen and Zhang, 1989; McCue and Dormand, 1993; Yoder, 1995; Correia and Laskar, 2001). Planetary perturbations can lead to large chaotic changes in the obliquity of Venus as a result of resonance overlap between the precession frequency and combinations of secular frequencies of changes in the orbital plane of Venus (Laskar and Robutel, 1993; Yoder, 1995). The current precession rate of Venus is about $45 \text{ arcsec yr}^{-1}$, corresponding to a precession period of about 28800 years (Yoder, 1997). If Venus initially had an obliquity lower than 90° and a rotation period shorter than 5 days, it could not have escaped the chaotic zone, which lies between 0° and 90° (Correia *et al.*, 2003). As a result, Venus could have reached the final retrograde state with obliquity $\epsilon = 180^\circ$ starting from any initial obliquity (Yoder, 1997; Correia and Laskar, 2001; Correia and Laskar, 2003a; Correia *et al.*, 2003). Planetary perturbations largely increase the probability of reaching a retrograde final state and make it the most probable final state for small initial rotation periods (< 5 days) or high initial obliquities (Correia and Laskar, 2001, 2003a; Correia *et al.*, 2003).

In the presence of planetary perturbations, capture into the 1:1 spin-orbit resonance becomes unlikely and the obliquity never goes to zero (Correia and Laskar, 2001, 2003a; Correia *et al.*, 2003). Small forced obliquity changes up to a few degrees are possible and are associated with small changes in the rotation rate (Yoder, 1995; Correia *et al.*, 2003). Therefore, with planetary perturbations included, the final states are no longer steady states. The current obliquity of Venus is 2.1° (Yoder, 1997). Planetary induced changes in the eccentricity of the orbit of Venus can also lead to long periodic changes in the rotation rate. Because the strengths of the tidal gravitational and thermal atmospheric torques have different dependencies on the radial distance from the Sun, changes in the eccentricity cause slight imbalances in these torques. The resulting rotation velocity changes can be up to a few kilometers per year at the equator, over time-scales of 10^6 years (Bills, 2005).

10.04.4.2 Wobble, LOD Variations, and Nutation

A departure of the mean position of the polar principal moment of inertia axis of Venus from the rotation axis has been determined from coherent two-way radio-tracking data of the Pioneer Venus Orbiter (PVO, 1978–1992) and Magellan (1990–1994). First results suggested a large orientation difference of a few degrees (Mottinger *et al.*, 1985; Bills *et al.*, 1987), much larger than the wobble amplitudes for the Earth and Mars, which are below an arcsec (Munk and MacDonald, 1960; Section 10.04.3). Later studies including Magellan data gave estimated values of about 0.5° (Konopliv *et al.*, 1993; McNamee *et al.*, 1993; Konopliv and Sjogren, 1994; Konopliv *et al.*, 1999), although Nerem *et al.* (1993, see also Konopliv *et al.*, 1993), who only used PVO data, obtained an even smaller amplitude of about 0.1° . Konopliv and Sjogren (1994) quote a formal error of 0.05° for the pole position.

Due to the slow rotation of Venus, $J_2 = (4.404 \pm 0.002) \times 10^{-6}$ (Konopliv *et al.*, 1999) is much smaller than for the Earth and not dominated by the contribution from rotation, which is on the order of 10^{-8} . The equatorial flattening is equally small, with $(B - A)/MR^2 = 4C_{22} = (1.57 \pm 0.02) \times 10^{-6}$. Because of these small flattenings and the long rotation period, the Chandler wobble has a very long period. In the rigid approximation, we have

$$P_{\text{CW}}^{\text{rigid}} = \frac{2\pi}{\Omega} \sqrt{\frac{AB}{(C-A)(C-B)}} \approx 50\,000 \text{ years} \quad [77]$$

(Mottinger *et al.*, 1985). Elasticity tends to increase this period by less than 1% ($k_2 = 0.295 \pm 0.066$; Konopliv and Yoder, 1996), whereas a liquid core would decrease the period by about 10%, depending on the size and density of the core (see eqn [46]). The relative effect of elasticity is much smaller than the almost 50% increase of the CW for Earth due to elasticity (Smith and Dahlen, 1981) because the polar flattening of Venus (or J_2) is mainly due to nonhydrostatic effects and not to rotational flattening as for the Earth. The deformation associated with the Chandler wobble, due to the differential rotation and the resulting centrifugal potential, then has a very small effect on the mean inertia moments and almost does not change the effective flattening. In eqn [46] for the Chandler wobble period, the

term k_2/κ is then very small compared to 1. The slow rotation of Venus also implies a much less efficient damping than on Earth (Yoder and Ward, 1979). A larger CW amplitude than for the Earth would then be possible even for a smaller excitation rate than that required to maintain the Earth's CW. Excitation could be due to seismic activity, mantle convection, or atmosphere/climate variations. By assuming quake-induced ΔC_{21} comparable to the Earth, Yoder and Ward (1979) estimated that a wobble amplitude $\leq 1^\circ$ requires a seismic acitivity rate on Venus $\leq 10^{-2}/Q(\sigma_{\text{CW}})$ times the rate of seismic activity on the Earth. The small observed wobble suggests that Venus is significantly less active than the Earth (Yoder, 1995, 1997).

Short periodic LOD variations and orientation changes in space (nutation) have not yet been observed and are expected to be small. Because of the almost perpendicular position of Venus' rotation axis with respect to the orbital plane (see Table 1), Venus does not have seasons as on the Earth and Mars, where the seasonal atmospheric cycle is the main driver of LOD variations. Periodic LOD variations of Venus can theoretically be calculated from general circulation models (GCM) of the atmosphere of Venus, but have, up to our knowledge, not yet been performed, most likely because the effects are small and the GCMs are not sufficiently precise. Better models and reliable ΔLOD estimates will hopefully become available with Venus Express data on the (low) atmosphere.

Nutations on Venus are expected to be somewhat smaller than on Earth. The solar torque, which causes the nutations, is much smaller than for the Earth mainly because the polar flattening is a factor 250 smaller (Konopliv and Yoder, 1996). The differences with respect to the Earth due to Venus' small obliquity and shorter distance to the Sun are much less significant. The shorter distance to the Sun would

tend to increase the nutations by a factor 2.6, and the small obliquity (see the $\sin 2\delta$ in eqns [64] and [65]) reduces the nutation amplitudes with respect to the Earth by a factor of 8. On the other hand, the slow rotation of Venus implies a weaker rotational stability and orientation changes such as nutation are larger for a given torque (see also eqns [62] and [63]), which almost compensates for the smaller torque. The small eccentricity of the orbit of Venus suggests that nutations other than the semi-annual nutation would be smaller.

For the Earth and Mars, nutation observations have and can be used to determine properties of the deep interior of these planets. For Venus, however, nutation observations have less geophysical potential because resonance effects with the FCN are expected to be negligible. The FCN period is about 10^5 yrs in inertial space (we assumed a similar core–mantle boundary flattening as for the total planet and neglected the very small effect of deformation), far from the main nutation excitation periods, which are associated with Venus' orbital motion.

10.04.5 Mercury

10.04.5.1 Long-Term Spin Evolution

10.04.5.1.1 Spin-orbit resonance

Based on radar observations of the rotation of Mercury made at the Arecibo Observatory in Puerto Rico (Pettengill and Dyce, 1965), Colombo (1965) proposed the idea that Mercury has a stable rotation period of exactly two-thirds of the orbital period. This 3:2 spin–orbit resonance is now universally accepted. Before, it was thought that Mercury's spin was in a 1:1 resonance with its orbital motion, just like the Moon in its orbit around the Earth (see, e.g., Chapman, 1988).

Table 1 Mass M , mean radius R , semi-major axis a , eccentricity e , obliquity ϵ , siderial rotation period P_{rot} , and orbital period P_{orb} for the four terrestrial planets

Property	Mercury	Venus	Earth	Mars
$M(10^{24} \text{ kg})$	0.3302	4.8685	5.9736	0.6419
$R(\text{km})$	2439	6052	6371	3390
$a(\text{AU})$	0.38710	0.72333	1.00000	1.52366
e	0.20563	0.00677	0.01671	0.09341
$\epsilon(\text{degree})$	< 1	177.9	23.45	25.19
$P_{\text{rot}}(\text{d})$	58.646	243.020	0.997	1.026
$P_{\text{orb}}(\text{d})$	87.969	224.695	365.256	686.980

Orbital elements are for J2000. See Sections 10.04.3–10.04.5.

Because of Mercury's closeness to the Sun, tidal friction can slow down the rotation of Mercury to rotation periods commensurate with the orbital period on a timescale much smaller than the age of the solar system (see eqn [57]). Correia and Laskar (2004) estimated that the time needed to despin Mercury from an initial rotation period of 10 h would be about 300 million years for a Q -value of 50. The final equilibrium rotation period for Mercury is different from the rotation synchronous with the orbital motion expected for tidal despining of planets in circular orbits because of Mercury's substantial eccentricity. If the rotation was synchronous, the angular rotation rate would be smaller than the orbital velocity close to perihelion and larger near to aphelion, resulting in accelerating and decelerating tidal torques, respectively. Because of its distance dependence, the tidal torque is larger near perihelion than at aphelion, and the mean tidal torque would act to spin up Mercury. For the present eccentricity ($e=0.206$) the equilibrium rotation velocity is $\omega=1.25685n$, and the present mean rotation rate of Mercury ($3n/2$) corresponds to the tidal equilibrium rotation rate for eccentricity $e_{3/2}=0.284927$ (Correia and Laskar (2004), who assume the averaged tidal torque to be linear in the tidal frequency). The tidal torque equilibrium is currently not satisfied because Mercury is locked in the 3:2 spin-orbit resonance.

Here we assume zero obliquity and principal axis rotation for the study of Mercury's rotation rate. Only the z -component of the Liouville equations must then be considered. For an entirely solid and rigid planet, eqn [24] reduces to

$$C\Omega\dot{m}_z = L_p \quad [78]$$

where

$$L_p = \frac{3}{2}(B-A)\frac{GM_\odot}{r^3}\sin 2\xi \quad [79]$$

is the solar torque on Mercury's permanent figure. We for the moment neglect the tidal torques. The distance r is the distance between the mass centers of the Sun and Mercury, and ξ is the angle between the direction of Mercury's axis of smallest moment of inertia (also called the long axis) and the direction to the Sun:

$$\xi = f - \phi \quad [80]$$

where f is the true anomaly and ϕ the rotation angle between the long axis of Mercury and the major axis of Mercury's orbit considered fixed (see Figure 4).

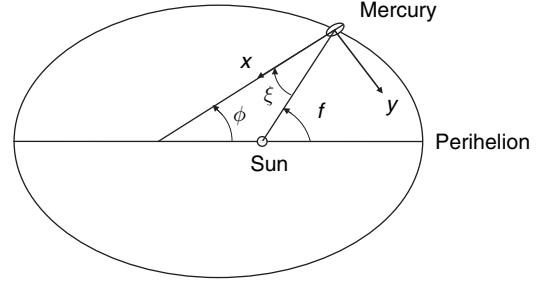


Figure 4 Geometry of Mercury's spin-orbit motion. Mercury's long axis makes an angle ξ with respect to the Sun and an angle ϕ with respect to an inertial reference axis, here taken to be temporarily coinciding with the semimajor axis of Mercury's orbital motion around the Sun.

We consider a rotation close to the 3:2 spin-orbit resonance and introduce the angle $\gamma = \phi - (3/2)M$, where M is the mean anomaly. At perihelion passage, γ is the angle between the direction of Mercury's long axis and the direction to the Sun. From the definitions of the angles γ and ξ , we have

$$\xi = f - \gamma - 3/2M \quad [81]$$

and the torque equation can be expressed as

$$C\ddot{\gamma} = -\frac{3}{2}n^2(B-A)\frac{a^3}{r^3}\sin(2\gamma + 3M - 2f) \quad [82]$$

where the mean motion

$$n = \sqrt{G(M_\odot + M_{\text{Merc}})/a^3} \approx \sqrt{GM_\odot/a^3}$$

is used. The time-dependent factor $(1/r^3)\sin(2\gamma + 3M - 2f)$ of the torque can be expanded in terms of the mean anomaly M as

$$\begin{aligned} & \frac{1}{r^3}\sin(2\gamma + 3M - 2f) \\ &= \frac{1}{a^3} \sum_q G_{20q}(e)\sin[2\gamma + (1-q)M] \end{aligned} \quad [83]$$

(e.g., Murray and Dermott, 1999), where $G_{20q}(e)$ are eccentricity functions (Kaula, 2000). By developing the sine terms and retaining only terms up to e^4 in the eccentricity functions, we have

$$\begin{aligned} & \frac{d^2\gamma}{dM^2} + \frac{3}{4}\frac{B-A}{C}e\left(7 - \frac{123}{8}e^2\right)\sin 2\gamma \\ &= -\frac{3}{2}\frac{B-A}{C}\left[\left(1 - 11e^2 + \frac{959}{48}e^4 + \dots\right)\sin M \right. \\ & \quad \left. - \frac{1}{2}e\left(1 + \frac{421}{24}e^2 + \dots\right)\sin 2M + \dots\right] + \dots \end{aligned} \quad [84]$$

(Peale 2005). The angle γ is considered very small, and time-dependent forcing terms proportional to $\sin 2\gamma$ have been neglected compared to the other terms in the right-hand side.

For the study of the stability of the 3:2 resonance, eqn [84] is averaged over an orbital period. Since the torques are small, ξ does not change much over an orbital period, and ξ (or $\dot{\gamma}$) can be kept constant in the averaging procedure. The periodic forcing terms in the right-hand member of eqn [84] then vanish, and the equation reduces to the harmonic oscillator (or pendulum) equation

$$\frac{d^2\gamma}{dM^2} + \frac{3}{4} \frac{B-A}{C} e \left(7 - \frac{123}{8} e^2 \right) \sin 2\gamma = 0 \quad [85]$$

which has $\gamma=0$ as a stable solution. This solution corresponds to the long axis of Mercury pointing toward the Sun at perihelion.

The effect of tidal dissipation can be included by adding the averaged tidal torque $\langle L_T \rangle$ divided by $n^2 C$ to the right-hand side of eqn [85]. If $\langle L_T \rangle$ can be assumed to be constant (independent of γ , $\dot{\gamma}$, t), the stability of the 3:2 resonance is maintained if the averaged tidal torque does not exceed the maximum possible restoring torque:

$$|\langle L_T \rangle| < \frac{3}{4} (B-A) n^2 e \left(7 - \frac{123}{8} e^2 \right) \quad [86]$$

(Colombo and Shapiro, 1966; Goldreich and Peale, 1966b). By using eqn [55], this translates into a condition on the permanent deformation and the tidal dissipation factor Q :

$$\frac{B-A}{4MR^2} > \frac{k_2}{14Q} \left(\frac{R}{a} \right)^3 \frac{M_\odot}{M} \frac{1 + \frac{123}{56} e^2}{e} \approx \frac{7 \times 10^{-8}}{Q} \quad [87]$$

where we used $k_2 \approx 0.4$ (Spohn *et al.*, 2001; Van Hoolst and Jacobs, 2003). Given that the moment of inertia difference $B-A=(4\pm 2)\times 10^{-5} MR^2$ (Anderson *et al.*, 1987), and that Q is in the range from 10 to 500 (Goldreich and Soter, 1966), this stability condition is certainly satisfied. For a circular orbit, the 3:2 resonance would not be stable. Note that a large equatorial flattening and a large eccentricity contribute to the stability of the resonance. Several other resonances in the vicinity of the 3:2 resonance are also stable (Goldreich and Peale, 1966b). Why then did Mercury select the 3:2 resonance?

A necessary condition for capture is that the averaged tidal torque is not symmetric about $\dot{\gamma}=0$, otherwise a decelerating Mercury would continue to

decrease its rotation rate (Goldreich and Peale, 1966b; Peale, 1976a). If the phase lag is proportional to the tidal frequency (Q inversely proportional to tidal frequency), this condition is satisfied. Even for constant Q dissipation models, the torque depends on the sign of the phase angle (or the sign of $\dot{\gamma}$; e.g., Kaula, 1964), and it can be concluded that the averaged tidal torque is expected to be asymmetric about $\dot{\gamma}$ for all tidal models (Goldreich and Peale, 1966b). The precise capture probability is quite model dependent. Assuming the dissipation factor to be inversely proportional to the frequency, Goldreich and Peale (1966b) found that Mercury has the largest probability to be captured in the 3:2 resonance, with a probability of 6.7%. The capture probability is rather small, but can be increased if dissipative core–mantle coupling is included. Other resonances will then also have increased capture probabilities, and specific core viscosity values are needed to avoid capture in the 2:1 resonance (Peale and Boss, 1977). In these calculations, the present value of the eccentricity was used. However, Mercury's eccentricity shows large chaotic variations in time between zero and about 0.5 due to planetary influences (Laskar, 1994). As a result, over time, Mercury's eccentricity could have passed several times through the value $e_{3/2}$ at which tidal torques tend to drive Mercury's rotation velocity to the 3:2 resonant value, thereby considerably increasing the capture probability (Correia and Laskar, 2004). Moreover, Mercury could have escaped from an initial resonance capture when the eccentricity decreased to low values close to zero. Correia and Laskar (2004) numerically followed the evolution of Mercury for 1000 different but close initial conditions and found that the 3:2 spin–orbit resonance is the most likely final state, with a probability of 55.4%.

10.04.5.1.2 Precession about the Cassini state and nutation

Mercury is thought to occupy the Cassini state 1, in which the rotation axis and the orbit normal remain coplanar with the normal to the Laplace plane as the rotation axis of Mercury precesses about the orbit normal and the orbit normal precesses about the normal to the Laplace plane (Colombo, 1966; Peale, 1969; Ward, 1975). The Laplace plane is defined as the plane about which Mercury's orbit precesses with constant inclination between the two planes. A Cassini state represents an extreme (equilibrium point) in the Hamiltonian of motion, made integrable by limiting the planetary perturbations (mainly due to Venus, the Earth, and Jupiter) to a single term

corresponding to a regular motion of the ascending node, and has first been determined empirically for the Moon (Cassini, 1693). Tidal friction drives the spin of Mercury to Cassini state 1 from almost any initial condition on a timescale that is short compared to the age of the solar system (Peale, 1974). The primordial obliquity of Mercury was probably below 90° , otherwise core–mantle friction could have driven Mercury to the Cassini state 3, similar to Cassini state 1 but with an obliquity 180° larger (Peale, 1976a). Provided that Mercury was born as a fast rotator, it must have passed a large chaotic zone in which the obliquity shows large and chaotic variations of tens of degrees within a few million years (Laskar and Robutel, 1993). The obliquity of the final evolved Cassini state 1 is given by

$$\frac{C}{MR^2} = \frac{\left[\mathcal{J}_2/(1-e^2)^{3/2} + eC_{22}\left(7 - \frac{123}{8}e^2\right) \right] n}{(\sin I/\epsilon_C) - \cos I} \mu \quad [88]$$

(Peale, 1969), where ϵ_C is the obliquity of Cassini state 1, I the inclination of the orbital plane to the Laplace plane, and μ the precession rate of Mercury's orbit about the normal to the Laplace plane. Because the rotation axis is coplanar with the orbit normal and the normal to the Laplace plane in a Cassini state, the obliquity of the rotation axis is constant. For the Mariner 10 gravitational coefficients ($\mathcal{J}_2 = (6.0 \pm 2.0) \times 10^{-5}$ and $C_{22} = (1.0 \pm 0.5) \times 10^{-5}$; Anderson *et al.*, 1987) and $C/MR^2 \approx 0.34$, the theoretical mean obliquity is equal to 1.6 arcmin, as can be determined from eqn [88] (Peale, 1988) and has also been obtained by Rambaux and Bois (2004) by means of a numerical integration of the full orbital and rotational motion of Mercury in the solar system. This value is below the measurement accuracy of the obliquity determinations from Mariner 10 observations (Klaasen *et al.*, 1976) and Earth-based radar measurements (Anderson *et al.*, 1996). This situation is likely to change in the near future. The most promising Earth-based technique to determine Mercury's obliquity is called Radar Speckle Displacement Interferometry (RSDI), which consists of the determination of the epoch and time delay that maximize the cross-correlation of the back-scattered radar signal at two antennas separated by a long baseline. It would allow determining Mercury's obliquity with a precision of a few arcseconds (Holin, 1999). The theoretical mean obliquity is also within reach of the MESSENGER mission (Solomon *et al.*, 2001), and the BepiColombo mission is aiming for an arcsecond precision (Milani *et al.*, 2001).

The precession period of both the orbit normal and the rotation axis around the normal to the Laplace plane is about 280 000 year (Rambaux and Bois, 2004). Besides this slow precession, the rotation axis of Mercury also precesses about its position in the Cassini state if Mercury does not occupy but is close to the Cassini state (Peale, 1974; Ward, 1975; Rambaux and Bois, 2004). If we neglect the much slower orbital precession, the latter precession due to the solar torque (also called free precession), can be compared to the precession of the Earth about the normal to the ecliptic. However, due to Mercury's 3:2 spin–orbit resonance, the classical analytical expression for the precession rate (see, e.g., Smart, 1953) must be extended with a term in the equatorial flattening (or C_{22}). By averaging the Liouville equations without deformation effects over the orbital period, the precession period can be derived, correct up to order e^3 as

$$P_{\text{prec}} = \frac{3\pi C'''}{nMR^2} \left\{ \left[\frac{3\mathcal{J}_2}{2(1-e^2)^{3/2}} + \frac{3}{2}eC_{22}\left(7 - \frac{123}{8}e^2\right) \right]^2 - C_{22}^2 \frac{159}{16}e^6 \right\}^{-1/2} \quad [89]$$

(Peale, 2005). For a solid core, $C = 0.34MR^2$, and the gravitational coefficients of Anderson *et al.* (1987), the precession period is 1062 years. This agrees very well with the period of 1066.91 years obtained with the full numerical method of Rambaux and Bois (2004) and with 1065.08 years derived in an Hamiltonian theory (D'Hoedt and Lemaître, 2005; see also Yseboodt and Margot, 2006). The period is proportional to C''' , so the precession period decreases with larger liquid core (larger core moment of inertia), as expected for short periodic mantle precession. The C_{22} terms decrease the traditional precession period by nearly 30% and cause the precession motion to be slightly elliptical (Peale, 2005).

Tidal friction and dissipative core–mantle coupling will damp free precession, that is, will reduce the obliquity variations, and will drive the rotation axis to the Cassini state if the orbital inclination to the invariable plane is fixed and Mercury precesses about the Cassini state at constant rate (Peale, 1974; Ward, 1975; Peale, 2005). Because of planetary perturbations, the precession rate and inclination show large variations, which suggests that Mercury cannot remain in the Cassini state. However, the precession period is much smaller than the shortest timescales of about 10^4 to 10^6 years for the variations in the

planetary orbital parameters. Therefore, the rotation axis accurately tracks the changes in the orientation of the orbit normal. If Mercury has once been driven close to the Cassini state by dissipative effects, it is expected to remain very close to the instantaneous location of the Cassini state (Goldreich and Toomre, 1969; Peale, 1974). With a typical core viscosity value for the Earth's core determined from *ab initio* calculations (de Wijs *et al.*, 1998) and $Q \approx 100$, Peale (2005) obtained free precession damping timescales of about 10^5 years. This is much shorter than the age of the solar system, suggesting that free precession is either completely damped or recently excited (Peale, 1974; Ward, 1975). On the other hand, the damping timescale is somewhat longer than the 10^4 year timescale of the Laplace plane reorientation, and the changing Cassini state may never be reached. Because the changes in the Cassini state are small and the damping timescale is only slightly longer than the orbital variations, Yseboodt and Margot (2006) argue that the Cassini state has been reached.

In their numerical integration of Mercury's spin-orbit motion in the solar system, Rambaux and Bois (2004) obtained obliquity variations with the free precession period of a few arcmin around a mean value of 1.6 arcmin. These variations depend on the initial conditions and can be strongly reduced by using suitable initial conditions close to the Cassini state. For initial conditions corresponding to the Cassini state, Yseboodt and Margot (2006) found that planetary perturbations do not excite free precession and the rotation axis remains in the Cassini state, even if the Cassini obliquity changes due to the precession of the orbit. In this study, the integration is over a few thousand years, and only the secular parts of the perturbing potential for seven exterior planets is included in the Hamiltonian formalism. On this relatively short timescale, the changes in Cassini state due to the full planetary perturbation is very small. Over a 12 kyr interval centered on J2000, the Cassini state obliquity (eqn [88]) increases from 1.64 to 1.73 arcmin (Yseboodt and Margot, 2006). Peale (2006) followed the spin position and the Cassini state position during short timescale orbital variations over 20 000 years as well as during long timescale variations over the past 3 My, and showed that the spin axis remains within one arcsec of the Cassini state if it initially occupied the Cassini state. On the other hand, Bills and Comstock (2005) obtained obliquity variations forced by the orbital motion on timescales of Myrs with typical obliquity values of 2–4 arcmin, values that are larger than the

Cassini state obliquities determined from eqn [88]. The obliquity history looks very much like a scaled version of the inclination history, with the obliquity variations being roughly 200 times smaller than the corresponding inclination variations and typical periods of about 100 kyr and 1 My (Bills and Comstock, 2005).

Another source of deviations from the Cassini state are short periodic variations in obliquity and longitude caused by the solar torque. These variations, similar to the nutations for the Earth and Mars, have main frequencies equal to the harmonics of the orbital frequency (Carpentier and Roosbeek, 2003). The amplitudes are on the order of 0.1 arcsec or smaller (Carpentier and Roosbeek, 2003; Rambaux and Bois, 2004). They have been calculated for a rigid Mercury, but are not expected to be largely amplified when the interior properties of Mercury are taken into account. For the Earth and Mars, the FCN frequency is close to the main nutation frequencies, and some nutations are resonantly amplified. For Mercury, the FCN period is about 500 years, as can be estimated from eqn [47], and is far from the period of the nutations.

10.04.5.2 Librations

10.04.5.2.1 Free and forced libration

Because of the gravitational torque on the permanent figure of Mercury in its elliptical orbit, the rotation of Mercury is not constant and varies about a mean state. For the analysis of these librations, we include the liquid core, but still neglect deformation (no incremental inertia terms). Liouville equations [24] and [36] can then be written as

$$C\Omega\dot{m}_z + C^f\Omega m_z^f = \frac{3}{2}(B-A)\frac{GM_{\odot}}{r^3}\sin 2\xi \quad [90]$$

$$C^f\Omega(\dot{m}_z + \dot{m}_z^f) = 0 \quad [91]$$

where the last equation indicates that the core does not follow the rotation variations, or librations, of the mantle. By substituting eqn [91] into eqn [90], and taking into account that $C = C^m + C^f$, the Liouville equation for the whole planet can be transformed to a Liouville equation for the mantle:

$$C^m\ddot{\gamma} = \frac{3}{2}(B-A)\frac{GM_{\odot}}{r^3}\sin 2\xi \quad [92]$$

where the libration γ is now defined in terms of the direction of the long axis of the mantle. With

inclusion of a liquid core, the only change in eqns [82] and [84] is that C must be replaced by C'' .

Equation [84] with C replaced by C'' is an equation for a forced harmonic oscillator describing forced libration in longitude. The angle γ is usually called the libration angle. The forcing is at 1 orbital period, 1/2 orbital period, and higher subharmonics of the orbital period, which have been neglected here. Due to the small ratio $(B - A)/C''$, the amplitude γ_1 of the 88 day libration is equal to the amplitude of the $\sin M$ term in eqn [84]:

$$\gamma_1 = \frac{3(B - A)}{2C''} \left(1 - 11e^2 + \frac{959}{48}e^4 + \dots \right) \quad [93]$$

(Peale, 1972), and the amplitude of the $\sin 2M$ term is about 11% of that (Jehn *et al.*, 2004).

The libration eqn [85] averaged over an orbital period governs free libration. When the long axis does not point toward the Sun at perihelion, the averaged gravitational torque on the permanent figure of Mercury tends to restore the alignment, and the long axis will librate around the direction to the Sun at perihelion. The oscillation period is given by

$$P_{\text{free}} = \frac{2\pi}{n} \left[\frac{1}{3} \frac{C''}{B - A} e \left(\frac{7}{2} - \frac{123}{16} e^2 \right) \right]^{1/2} \quad [94]$$

Evidently, free libration is also due to the gravitational interaction between the Sun and Mercury, as is forced libration. The distinctive feature of free libration as opposed to forced libration is that the amplitude and phase of free libration can only be determined theoretically if its excitation and dissipation are well modeled.

10.04.5.2.2 Forced libration and interior structure

The interior of Mercury is not well known, and libration offers a unique opportunity to study some important aspects of it. Both the amplitude of the forced libration at 88 days and the free libration period depend on the geophysically interesting mantle moment of inertia, which is related to the size and density distribution of the mantle. Mantle must be understood here as the solid part of the planet above the liquid core that partakes in the libration, and thus also includes the crust. If there is no liquid core, the moments of inertia referring to the mantle are the total planetary moments of inertia. The magnetic field observed by Mariner 10, if caused by dynamo action, is evidence for a liquid outer core, which, according to thermal evolution models, very likely

contains a solid inner core (Stevenson *et al.*, 1983; Schubert *et al.*, 1988; Spohn, 1991). The inner core is possibly even relatively large (Stevenson *et al.*, 1983; Hauck *et al.*, 2004).

The large mean density of Mercury points towards a large core, which is thought to consist of iron with only limited additions of lighter elements. Sulfur is considered as the most likely candidate (Schubert *et al.*, 1988; Harder and Schubert, 2001; Spohn *et al.*, 2001; Van Hoolst and Jacobs, 2003). The lack of data and Mercury's unknown origin make it impossible to constrain the composition of the core. The sulfur (and other light element) concentration depends critically on where Mercury accreted in the initial solar system and on where the planetesimals that formed the planet originated from (Wetherill, 1988). If Mercury formed close to its present position, its sulfur concentration is probably very low (Lewis, 1988). More likely however, Mercury formed in the same feeding zones as Earth, Venus, and Mars, and its light element concentration could be higher and closer to that of those planets. The Earth has a light element concentration of about 10 wt.% (Poirier, 1994), and for Mars a core sulfur concentration of 14 wt.% is often considered (Deibus and Wänke, 1985; Longhi *et al.*, 1992).

Figure 5 shows the forced libration amplitude for several series of interior structure models. The models have a core with a composition ranging from almost pure iron to an iron–sulfur assemblage with 14 wt.% sulfur, and an inner core radius anywhere between 0 km (entirely liquid core) and the radius of the core (entirely solid core) (Van Hoolst and Jacobs, 2003). The forced libration amplitude is calculated by using eqn [93] for the moment of inertia difference

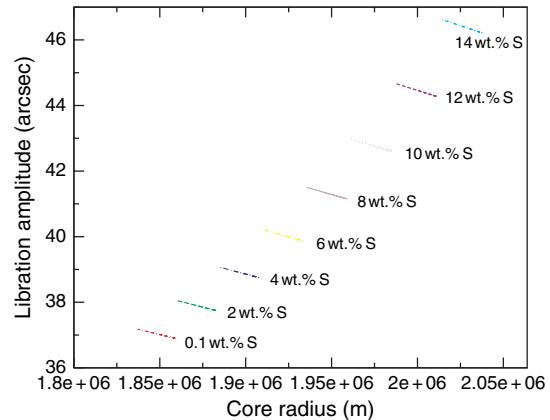


Figure 5 Forced libration amplitude as a function of core radius.

$B - A = 4 \times 10^{-5} M R^2$ (Anderson *et al.*, 1987). The amplitude is about 21 arcsec for models with a solid core, and between 37 and 47 arcsec for models with a liquid core. For larger cores and smaller mantle moments of inertia, the libration amplitude increases since the libration is essentially a libration of the mantle. The models have the same mantle density and, as their mass and radius take the observed values, the sulfur concentration directly influences the size of the core and as such also the mantle moment of inertia. Therefore, the libration amplitude also increases with the mean sulfur concentration in the core. If the mantle density profile can be determined from other means, the observation of the libration amplitude strongly constrains the core composition. A further restriction is that the moment of inertia difference $B - A$ must be known, which can be derived with high accuracy from tracking a spacecraft orbiting about Mercury, such as MESSENGER or BepiColombo.

The libration amplitude is sufficiently large for being within reach of accurate ground-based radar experiments. With RSDI, the instantaneous rotation rate of Mercury, and thus the libration, can be determined from Earth with unprecedented accuracy of about 1 arcsec, according to theoretical estimates (Holin, 2001). Preliminary results of RSDI observations of Mercury with the radio telescopes at Goldstone, Green Bank, and Arecibo showed a large forced libration amplitude of about 60 ± 6 arcsec. This amplitude is larger than the model values calculated above with $B - A = 4 \times 10^{-5} M R^2$ (see eqn [93]), even for the largest liquid cores (see Figure 5). By taking into account that the equatorial moment of inertia difference $B - A$ has a 50% uncertainty (Anderson *et al.*, 1987), it can be concluded with very high probability that Mercury has a liquid core (Margot *et al.*, 2002, 2004).

Space missions offer the advantage that not only libration but also other quantities needed for the precise interpretation of the libration data can be measured. In particular, the accurate measurement of the four parameters γ_1 , \mathcal{J}_2 , C_{22} , and ϵ_C are sufficient to probe the interior of Mercury (Peale, 1976b). From these parameters the three factors in the equation

$$\left(\frac{C_m}{B - A} \right) \left(\frac{B - A}{M R^2} \right) \left(\frac{M R^2}{C} \right) = \frac{C_m}{C} \quad [95]$$

can be determined and a value for the ratio of the mantle to the total polar moment of inertia can be obtained. The first factor follows from the libration, the second from the gravity coefficient C_{22} , and the

third from the Cassini state obliquity (eqn [88]). The space missions MESSENGER and BepiColombo are expected to measure the degree-two gravity coefficients with an accuracy better than 1% and the obliquity and forced libration amplitude better than a few arcsecs (Solomon *et al.*, 2001; Milani *et al.*, 2001).

Since it is assumed in the derivation of the libration equations that the librations are with respect to the equilibrium state in which the core corotates with the mantle, the liquid core is tacitly assumed to follow the mantle during the long-period precession. Viscous core–mantle coupling is sufficiently large for this condition to be satisfied (Peale, 1988). Mercury is assumed to occupy Cassini state 1 in this procedure to determine C_m/C . According to eqn [88], the moment of inertia C is approximately inversely proportional to the obliquity ϵ_C of the Cassini state. The obliquity measurements will give the true obliquity of Mercury, which could be different from ϵ_C if orbital perturbations cause deviations from the Cassini state or if free precession is excited (nutations can be neglected). Excitation of free precession by impacts is very unlikely (Peale, 2005), and deviations due to planetary perturbations are most likely below the observational precision (Peale, 2006).

For a correct interpretation of libration observations, all possible effects on the rotation angle variations of Mercury have to be quantified, such as the effect of neglected core–mantle couplings in eqn [93]. These couplings have been shown to be important for changes in the rotation angle of the Earth. For example, on a decadal timescale, the changes in the rotation angle of the Earth are attributed to the exchange of angular momentum between the core and the solid mantle (Jault *et al.*, 1988), but there is no general agreement on which physical mechanism dominates.

Electromagnetic coupling between core and mantle arises from the flow of electric current in the weakly conducting mantle. The mantle conductivity is assumed to be concentrated in a thin layer at the CMB, possibly resulting from chemical reactions between the mantle silicates and the conducting liquid iron assemblage (Buffett *et al.*, 2000). Holme (1998) showed that electromagnetic coupling is capable of explaining the observed changes in the rotation rate of the Earth. Peale *et al.* (2002) applied Buffett's (1992) formalism for electromagnetic coupling to Mercury and estimated the torque related to the poloidal magnetic field from Mercury's observed magnetic dipole moment of $6 \times 10^{12} \text{ Tm}^3$ (Russell *et al.*, 1988). The time constant for the decay of the

relative angular velocity between core and mantle due to the electromagnetic torque is estimated to be on the order of 10^6 years. As this is much larger than the 88 day period of the forced libration, the electromagnetic torque does not affect the forced libration (Peale *et al.*, 2002). The torque could, however, be larger, and the decay time smaller, since only the observed dipole moment is used in the calculation and the torque involving the toroidal magnetic field has been neglected. Possibly, the dipole field of Mercury is even very small compared to the total magnetic field in the core. Stanley *et al.* (2005) showed that thin shell dynamos, which could be applicable to Mercury, can produce magnetic fields of different geometry than the Earth's magnetic field and with Mercury-like dipolar field intensities. The toroidal field strength can be 100 times larger than the dipolar strength, and the poloidal field strength can be 10 times larger than the dipolar strength. Projecting the dipole field observed by Mariner 10 to the CMB then does not provide a good estimate for the CMB field.

In our derivation of eqn [93], we only considered inertial core–mantle coupling in the lowest-order approximation. In that approximation, the core does not follow the mantle libration, as would be the case for a symmetric core ($A^f = B^f$) that is not coupled to the mantle. If the core is not axially symmetric, it will in principle be coupled to the mantle, although the coupling is very small. Two different couplings have to be considered: (1) gravitational coupling between a nonaxially symmetric inner core and the mantle and (2) pressure and gravitational coupling between the nonaxisymmetric outer core and the mantle. To study the effect of the former coupling, Peale *et al.* (2002) numerically integrated the Liouville equations for the mantle and the inner core, with gravitational coupling according to Szeto and Xu's (1997) formalism included. The amplitude of the 88 day mantle libration with gravitational coupling is nearly equal to that of eqn [93] without gravitational coupling, even for a large inner core with a radius of up to 90% of the core radius and a large inner core flattening $(B^{ic} - A^{ic})/C^{ic}$ equal to the whole planet value $(B - A)/C$ (Peale *et al.*, 2002).

Pressure coupling between the liquid core and the mantle requires a nonspherical interface. If Mercury's mantle convection ceased long time ago (Reese and Peterson, 2002), the mantle is probably in hydrostatic equilibrium and the core–mantle boundary is axially symmetric: $A^f = B^f$. Mantle convection (e.g., Schubert *et al.*, 1988) on the other hand induces

topography on the CMB. In the thin Mercurian mantle (see Chapter 10.5), convection cells with relatively small horizontal length scales comparable to the mantle depth become unstable for the lowest Rayleigh numbers (Schubert *et al.*, 2001). The resulting deviations from the hydrostatic equilibrium form are not expected to contribute largely to $B - A$. Models for convection in homogeneous spheres indicate that one order of magnitude larger mantle Rayleigh numbers are needed for larger-scale convection cells to become unstable (Schubert *et al.*, 2001). Conzelmann and Spohn (1999) studied mantle convection in Mercury with a sophisticated numerical model of mantle convection for various mantle rheologies. For rheologies comparable to the Earth's mantle, their numerical models show that convection is organized in convection cells of large horizontal scale, which could contribute to an appreciable $B^f - A^f$. In the assumption of a homogeneous core liquid, a triaxial ellipsoidal liquid core and the mantle are coupled due to the pressure of the liquid on the ellipsoidal CMB (inertial coupling). Rambaux *et al.* (2007) have included the Liouville equations for the core and mantle for this pressure coupling in an accurate numerical model for the libration of Mercury. The difference with the amplitude of eqn [93] for the 88 day libration in longitude is on the order of only 0.01 arcsec, even for large flattenings up to 10^{-3} (Rambaux *et al.*, 2007).

Several authors have argued that pressure forces on the CMB small-scale topography (as opposed to the large-scale triaxiality) are likely the dominant cause of the observed decadal variations in the LOD of the Earth (Jault *et al.*, 1988; Hide, 1989), although others have questioned the effectiveness of this mechanism (Kuang and Bloxham, 1997). By rescaling the topographic torque on the Earth's mantle needed to explain the Earth's LOD variations to Mercury, the change in the forced libration can be shown to be essentially undetectable (Peale *et al.*, 2002). As a conclusion, the studied core–mantle couplings are too small to appreciably affect the amplitude of the 88 day forced libration in longitude, and the core can be assumed not to follow the mantle during the 88 day libration.

10.04.5.2.3 Free libration and interior structure

The free libration period given by expression [94] for an entirely solid Mercury is equal to 15.830 years, very close to the period of 15.847 years, obtained with a numerical model (Rambaux and Bois, 2004),

and to 15.857 years determined with a Hamiltonian approach of Mercury's rotational motion (D'Hoedt and Lemaître, 2005). As before, values $B - A = 4 \times 10^{-5} M R^2$ and $C = 0.34 M R^2$ were used here. The period depends strongly on the interior structure of Mercury. For the models with a liquid core used in Figure 5 and the same value for $B - A$, the free libration period is between about 10.5 and 12 years, much shorter than the period for an entirely solid model (Rambaux *et al.*, 2007). Since the free libration is essentially a mantle libration, the period is longer for larger moment of inertia of the mantle. The long period makes this effect more difficult to detect in a few orbital periods of Mercury, the expected lifetime of the space missions MESSENGER and BepiColombo. However, a combination of ground-based radar measurements, MESSENGER and BepiColombo would allow the period to be detected precisely and to place an additional constraint on the mantle moment of inertia C''' of Mercury.

The amplitude of the free libration depends on excitation and damping processes. Free libration will eventually be damped to zero in the absence of excitation. A main source of damping are the tides caused by the Sun. Mercury reacts viscoelastically to forces that tend to deform it, and inevitably some energy is lost. Most energy is probably dissipated in the inner core (Schubert *et al.*, 1988). A further potential source of damping is due to dissipative core–mantle coupling, either due to core viscosity or Ohmic dissipation in the core. The timescale for damping of the free libration by viscous core–mantle coupling and tides is on the order of 10^5 years (Peale, 2005). For the same dissipation parameters as used for the damping of the free precession, the viscous core–mantle coupling dominates the tidal damping by one order of magnitude (Peale, 2005). The damping timescales are short compared to the age of the solar system. Without recent excitation, the free libration is therefore expected to be completely damped. Impact excitation is very unlikely since the average time span between impacts of sufficient size is about 10^9 years with current cometary fluxes. A recent impact would have left an impact crater of at least 20 km diameter for the excitation of an observable amplitude of 0.1 amin (Peale, 2005). On the other hand, planetary perturbations provide a continuous source of excitation for the free libration as they change Mercury's orbit. For example, changes in the semi-major axis of Mercury's orbit imply a change in the mean motion, and cause the long axis

of Mercury (see libration angle γ) to be misaligned with respect to the direction to the Sun at pericenter. The amplitude is presently unknown. In future observations, free and forced librations are easily separable as they have different periods.

10.04.5.3 Wobble

Polar motion of Mercury has never been detected. Due to the very tenous atmosphere of Mercury, seasonal atmospheric excitation of polar motion, as on Earth, can be neglected. As for Venus, the Chandler wobble component of polar motion has a very long period for Mercury because of its slow rotation and small flattenings, which are two orders of magnitude larger than the polar rotational flattening. By neglecting the very small deformation effect due to the slow rotation (as for Venus), and assuming a rigid behavior, the CW or Euler period can be derived from eqn [46]. We have $P_{\text{CW}}^{\text{rigid}} \approx 965$ year (Rambaux and Bois, 2004) for $C = 0.34 M R^2$ and the Mariner 10 values of J_2 and C_{22} (Anderson *et al.*, 1987). With a large fluid core with moment of inertia half of that of the total planet moment of inertia, the Chandler wobble period is half the rigid value. No strong excitation is known at this timescale. Seismic activity is a potential source of CW excitation, but is expected to be orders of magnitude smaller than on Earth for lack of plate tectonics. On the other hand, dissipation of the CW is less efficient than on Earth due to the slow rotation of Mercury but, for weak excitation, will eventually damp the wobble to very small values.

10.04.6 Summary

The rotation of the four terrestrial planets of our solar system has changed substantially since the planets' formation. The rotation history cannot be uniquely described but recent advances have led to a deeper understanding of their current rotation state. All four planets could have experienced large chaotic variations in obliquity due to resonances between the precession rate and the orbit precession rate. Their obliquities cannot therefore be considered as primordial and almost do not constrain the primordial obliquities. The rotation rates have also been changed. Because Mars is farther from the Sun than the other terrestrial planets and has no large satellites, tidal braking has only had a small influence on its spin rate, which can be considered to be primordial.

Mercury and Venus are in a stable final rotation state, as a result of dissipative processes. Therefore, their current spin gives no insight into their primordial rotation. Although the occupied final spin states have been shown to be the most probable outcomes of the rotational evolution, the physical dissipation mechanisms are not well constrained and need further study.

Short-periodic rotation variations of the Earth have been studied since antiquity and are nowadays routinely tracked with submilliarcsecond accuracy (see Chapter 3.10). Theoretical techniques developed for the Earth are now being applied to the other terrestrial planets, but data have so far only been obtained for the largest rotation variation components with limited precision. Seasonal LOD variations on Mars have recently been determined with a precision of about 10% and can be used to constrain general circulation models of the Martian atmosphere and the seasonal sublimation/condensation cycle of the polar caps. The rotation rate variations determined from state-of-the art general circulation models differ by more than the observational uncertainty. Preliminary results for the libration of Mercury also have a precision of about 10% and show with very high probability that the core of Mercury is at least partially liquid.

The next decade promises to be an exciting one for the study of periodic rotations variations of Mars and Mercury. Continued radio-tracking observations of Mars orbiters such as Odyssey and MRO (Mars Reconnaissance Orbiter) will improve LOD variations and make it possible to resolve the Mars CW period. This will aid in the construction of accurate general circulation models and will constrain further the interior of Mars. With a small network of fixed landers with dual-band coherent radio transponders, often been proposed but still not definitely accepted by any of the space agencies, an order of magnitude better accuracies, on the order of a few mas, could be obtained. Besides improving LOD variations and polar motion, it would allow detecting the nonrigid contributions to nutations and so offer direct information on the core of Mars. Earth-based RSDI observations of Mercury and measurements by the MESSENGER and BepiColombo missions to Mercury will measure the libration and obliquity with an accuracy better than a few arcsec. The total planetary moment of inertia and the ratio of mantle to whole planet moment of inertia can then accurately be determined, with expected precisions of about 1% and 10%, respectively, yielding important

information on Mercury's deep interior. Moreover, Venus Express data are expected to better constrain the rotation variations induced by the atmosphere.

References

- Aharonson O, Zuber MT, Smith DE, Neumann GA, Feldman WC, andPrettyman TH (2004) Depth, distribution, and density of CO₂ deposition on Mars. *Journal of Geophysical Research* 109: E05004 (doi:10.1029/2003JE002223).
- Anderson DL (1989) *Theory of the Earth*. Boston, MA: Blackwell Scientific Publications.
- Anderson JD, Colombo G, Esposito PB, Lau EL, and Trager GB (1987) The mass, gravity field, and ephemeris of Mercury. *Icarus* 71: 337–349.
- Anderson JD, Jurgens RF, Lau EL, Slade MA, and Schubert G (1996) Shape and orientation of Mercury from radar ranging data. *Icarus* 124: 690–697.
- Andoyer H (1923) *Cours de Mécanique Céleste*. Paris: Gauthier-Villars.
- Barnes RT, Hide R, White HH, and Wilson CA (1983) Atmospheric angular momentum fluctuation, length-of-day changes and polar motion. *Proceedings of the Royal Society of London A* 387: 31–73.
- Barriot JP, Dehant V, Folkner W, et al. (2001) NEIGE: Net-lander ionosphere and geodesy experiment. *Advances in Space Research* 28(8): 1237–1249.
- Bills BG (2005) Variations in the rotation rate of Venus due to orbital eccentricity modulation of solar tidal torques. *Journal of Geophysical Research* 110: E11007 (doi:10.1029/2003JE002190).
- Bills BG and Comstock RL (2005) Forced obliquity variations of Mercury. *Journal of Geophysical Research* 110: E04006 (doi:10.1029/2003JE002116).
- Bills BG, Kieffer WS, and Jones RL (1987) Venus gravity: A harmonic analysis. *Journal of Geophysical Research* 92: 232–250.
- Bills BG, Neumann GA, Smith DE, and Zuber MT (2005) Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110: E07004 (doi:10.1029/2004JE002376).
- Bizouard C (1995) *Modélisation Astrométrique et Géophysique de la Rotation de la Terre*. PhD Thesis, Observatoire de Paris, France.
- Borderies N (1980) Theory of Mars rotation in Euler angles. *Astronomy and Astrophysics* 82: 129–141.
- Born JC and Born GH (1976) Secular acceleration of Phobos and Q of Mars. *Icarus* 27: 51–53.
- Bouquillon S and Souchay J (1999) Precise modeling of the precession-nutation of Mars. *Astronomy and Astrophysics* 345: 282–297.
- Bretagnon P and Francou G (1988) VSOP87: Planetary theories in rectangular and spherical variables. *Astronomy and Astrophysics* 202: 309–315.
- Breuer D and Spohn T (2003) Early plate tectonics versus single-plante tectonics on Mars: Evidence from magnetic field history and crust evolution. *Journal of Geophysical Research* 108(E7): 5072 (doi:10.1029/2002JE001999).
- Buffett BA (1992) constraints on magnetic energy and mantle conductivity from the forced nutations of the Earth. *Journal of Geophysical Research* 97(B13): 19581–19597.
- Buffett BA, Garnero EJ, and Jeanloz R (2000) Sediments at the top of the Earth's core. *Science* 290: 1338–1342.

- Buffett BA, Mathews PM, and Herring TA (2002) Modeling of nutation and precession: Effects of electromagnetic coupling. *Journal of Geophysical Research* 107: doi:10.1029/2000JB000056.
- Busse FH (1968) Steady fluid flow in a precessing spheroidal shell. *Journal of Fluid Mechanics* 33: 739–751.
- Carpenter RL (1964) Study of Venus by CW radar. *The Astronomical Journal* 69: 2–11.
- Carpenter RL (1966) Study of Venus by CW radar—1964 results. *The Astronomical Journal* 71: 142–152.
- Carpenter RL (1970) A radar determination of the rotation of Venus. *The Astronomical Journal* 75: 61–66.
- Carpentier G and Roosbeek F (2003) Analytical development of rigid Mercury nutation series. *Celestial Mechanics and Dynamic Astronomy* 86: 223–236.
- Cassini GD (1693) *Traité de l'origine e de Progrès de l'astronomie*. Paris.
- Cazenave A and Balmino G (1981) Meteorological effects on the seasonal variations of the rotation of Mars. *Geophysical Research Letters* 8: 245–248.
- Chao BF and Rubincam DP (1990) Variations of Mars gravitational field and rotation due to seasonal CO₂ exchange. *Journal of Geophysical Research* 95(B9): 14755–14760.
- Chapman CR (1988) Mercury: Introduction to an end-member planet. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 1–23. Tucson, AZ: University Arizona Press.
- Chapman S and Lindzen R (1970) *Atmospheric Tides. Thermal and Gravitational*. Dordrecht: Reidel.
- Chapront-Touzé M (1990) Orbits of the Martian satellites from ESAPHO and ESADE theories. *Astronomy and Astrophysics* 240: 159–172.
- Colombo G (1965) Rotational period of the planet Mercury. *Nature* 208: 575.
- Colombo G (1966) Cassini's second and third laws. *The Astronomical Journal* 71: 891–896.
- Colombo G and Shapiro II (1966) The rotation of the planet Mercury. *The Astronomical Journal* 145: 296–307.
- Conzelmann V and Spohn T (1999) New thermal evolution models suggesting a hot, partially molten Mercurian core. *Bulletin of the American Astronomical Society* 31: 1102.
- Correia ACM and Laskar J (2001) The four final rotation states of Venus. *Nature* 411: 767–770.
- Correia ACM and Laskar J (2003a) Long-term evolution of the spin of Venus II. Numerical simulations. *Icarus* 163: 24–45.
- Correia ACM and Laskar J (2003b) Different tidal torques on a planet with a dense atmosphere and consequences to the spin dynamics. *Journal of Geophysical Research* 108(E11): 5123 (doi:10.1029/2003JE002059).
- Correia ACM and Laskar J (2004) Mercury's capture into the 3/2 spin-orbit resonance as a result of its chaotic dynamics. *Nature* 429: 848–850.
- Correia ACM, Laskar J, and de Surgy ON (2003) Long-term evolution of the spin of Venus. I. Theory. *Icarus* 163: 1–23.
- Dahlen FA and Tromp J (1998) *Theoretical Global Seismology*. Princeton, NJ: Princeton University Press.
- Darwin GH (1908) *Tidal Friction and Cosmogony*, vol. 2, *Scientific papers*. New York: Cambridge University Press.
- Davies ME, Colvin TR, Rogers PG, et al. (1992) The rotation period, direction of the North Pole, and geodetic control network of Venus. *Journal of Geophysical Research* 97(E8): 13141–13151.
- Defraigne P, de Viron O, Dehant V, Van Hoolst T, and Hourdin F (2000) Mars rotation variations induced by atmosphere and ice caps. *Journal of Geophysical Research* 105(E10): 24563–24570.
- Defraigne P, Dehant V, and Pâquet P (1995) Link between the retrograde–prograde nutations and nutations in obliquity and longitude. *Celestial Mechanics and Dynamic Astronomy* 62: 363–376.
- Defraigne P, Dehant V, and Van Hoolst T (2001) Steady state convection in Mars' mantle. *Planetary and Space Science* 49: 501–509.
- Defraigne P, Rivoldini A, Van Hoolst T, and Dehant V (2003) Mars nutation resonance due to free inner core nutation. *Journal of Geophysical Research* 108(E12): 5128 (doi:10.1029/2003JE002145).
- Defraigne P and Smits I (1999) Length-of-day variations due to zonal tides for an inelastic earth in non-hydrostatic equilibrium. *Geophysical Journal International* 139: 563–572.
- Dehant V, de Viron O, and Greff-Lefftz M (2005) Atmospheric and oceanic excitation of the rotation of a three-layer Earth. *Astronomy and Astrophysics* 446: 345–355 (doi:10.1051/0004-6361:20053825).
- Dehant V, de Viron O, Karatekin Ö, and Van Hoolst T (2006) Excitation of Mars polar motion. *Astronomy and Astrophysics* 446: 345–355 (doi:10.1051/0004-6361:20053825).
- Dehant V, Defraigne P, and Van Hoolst T (2000a) Computation of Mars' transfer function for nutations, tides and surface loading. *Physics of the Earth and Planetary Interiors* 117: 385–395.
- Dehant V, Hinderer J, Legros H, and Lefftz M (1993) Analytical approach to the computation of the Earth, the outer core and the inner core rotational motions. *Physics of the Earth and Planetary Interiors* 76: 259–282.
- Dehant V, Van Hoolst T, and Defraigne P (2000b) Comparison between the nutations of the planet Mars and the nutations of the Earth. *Surveys in Geophysics* 21: 89–110.
- Dehant V, Van Hoolst T, de Viron O, Greff-Lefftz M, Legros H, and Defraigne P (2003) Can a solid inner core of Mars be detected from observations of polar motion and nutation of Mars? *Journal of Geophysical Research* 108(E12): 5127 (doi:10.1029/2003JE002140).
- Deleplace B and Cardin P (2006) Viscomagnetic torque at the core–mantle boundary. *Geophysical Journal International* 167: 557–566 (doi:10.1111/j.1365-246X.2006.03180.x).
- de Wijs GA, Kresse G, Vočadlo L, Dobson D, Alfé, Gillan MJ, and Price GD (1998) The viscosity of liquid iron at the physical conditions of the Earth's core. *Nature* 392: 805–807.
- D'Hoedt S and Lemaitre A (2005) The Spin-orbit resonance of Mercury: A Hamiltonian approach. In: Kurtz DW (ed.) *Transits of Venus: New Views of the Solar System and Galaxy, Proceedings of IAU Colloquium 196*, pp. 263–270. Cambridge: Cambridge University Press.
- Dobrovolskis AR (1980) Atmospheric tides and the rotation of Venus. II. Spin evolution. *Icarus* 41: 18–35.
- Dobrovolskis AR and Ingersoll AP (1980) Atmospheric tides and the rotation of Venus. I. Tidal theory and the balance of torques. *Icarus* 41: 1–17.
- Dones L and Tremaine S (1993) On the origin of planetary spins. *Icarus* 103: 67–92.
- Dreibus G and Wänke H (1985) Mars, a volatile-rich planet. *Meteoritics* 20: 367–381.
- Dyce RB, Pettengill GH, and Shapiro II (1967) Radar determination of the rotations of Venus and Mercury. *The Astronomical Journal* 72: 351.
- Folkner WM, Kahn RD, Preston RA, et al. (1997b) Mars dynamics from Earth-based tracking of the Mars pathfinder lander. *Journal of Geophysical Research* 102(E2): 4057–4064.
- Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997a) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars pathfinder. *Science* 278: 1749–1752.
- Forget F, Hourdin F, Fournier R, et al. (1999) Improved general circulation models of the Martian atmosphere from the

- surface to above 80 km. *Journal of Geophysical Research* 104(E10): 24155–24176.
- Gaetani GA and Grove TL (1997) Partitioning of moderately siderophile elements among olivine, silicate melt, and sulfide melt: Constraints on core formation in the Earth and Mars. *Geochimica et Cosmochimica Acta* 61: 1829–1846.
- Getino J and Ferrández JM (2000) Effects of dissipation and a liquid core on forced nutations in Hamiltonian theory. *Geophysical Journal International* 142(3): 703–715.
- Gold T and Soter S (1969) Atmospheric tides and the resonant rotation of Venus. *Icarus* 11: 356–366.
- Goldreich P and Peale SJ (1966b) Spin–orbit coupling in the solar system. *The Astronomical Journal* 71: 425–438.
- Goldreich P and Peale SJ (1966a) Resonant rotation for Venus? *Nature* 209: 1117–1118.
- Goldreich P and Peale SJ (1970) The obliquity of Venus. *The Astronomical Journal* 75: 273–284.
- Goldreich P and Soter S (1966) Q in the solar system. *Icarus* 5: 375–389.
- Goldreich P and Toomre A (1969) Some remarks on polar wandering. *Journal of Geophysical Research* 74: 2555.
- Goldstein H (1950) *Classical Mechanics*. Reading, MA: Addison-Wesley.
- Goldstein RM (1964) Venus characteristics by Earth-based radar. *The Astronomical Journal* 69: 12–18.
- Golombek MP (2002) A revision of Mars seismicity from surface faulting. *Lunar and Planetary Science XXXIII*.
- Golombek MP, Banerdt WB, Tanaka KL, and Tralli DM (1992) A prediction of Mars seismicity from surface faulting. *Science* 258: 979–981.
- Gross R (2000) The excitation of the Chandler wobble. *Geophysical Research Letters* 27: 2329–2332.
- Gudkova TV and Zharkov VN (2004) Mars: Interior structure and excitation of free oscillations. *Physics of the Earth and Planetary Interiors* 142: 1–22.
- Gwynn CR, Herring TA, and Shapiro I (1986) Geodesy by radio interferometry: Studies of the forced nutations of the Earth. 2. Interpretation. *Journal of Geophysical Research* 91(B5): 4755–4766.
- Harder H and Christensen UR (1996) A one-plume model of martian mantle convection. *Nature* 380: 507–509.
- Harder H and Schubert G (2001) Sulfur in Mercury's core? *Icarus* 151: 118–122.
- Harri AM, Marsal O, Lognonne P, et al. (1999) Network science landers for Mars. *Advances in Space Research* 23: 1915–1924.
- Hauck SA, II, Dombard AJ, Phillips RJ, and Solomon SC (2004) Internal and tectonic evolution of Mercury. *Earth and Planetary Science Letters* 222: 713–728.
- Heiskanen WA and Moritz H (1967) *Physical Geodesy*. San Francisco, CA: W. H. Freeman.
- Herschel W (1774) On the remarkable appearances of the polar regions of the planet Mars, the inclination of its axis, the position of its poles, and its spheroidal figure, with a few hints relating to its real diameter and atmosphere. *Philosophical Transactions* 24: 233–273.
- Hide R (1989) Fluctuations in the Earth's rotation and the topography of the core–mantle interphase. *Philosophical Transactions of the Royal Society of London A* 328: 351–363.
- Hilton JL (1991) The motion of Mars' pole. I: Rigid body precession and nutation. *The Astronomical Journal* 102: 1510–1527.
- Hilton JL (1992) The motion of Mars' pole. II The effect of an elastic mantle and a liquid core. *The Astronomical Journal* 103: 619–637.
- Hinderer J, Legros H, and Amalvict M (1982) A search for the Chandler and nearly diurnal free wobbles using Liouville equations. *Geophysical Journal of the Royal Astronomical Society* 71: 303–332.
- Holin I (1999) Very long baseline radar interferometry for investigation of Mercury's spin and interior (in Russian). *Uspekhi Sovremennoi Radioelektroniki* 7: 16–28.
- Holin I (2001) Precise measurements of Mercury's obliquity and wobble by Earth-based radar speckle displacement interferometry. Paper presented at the Vernadsky Institute-Brown University Microsymposium 34, Topics in Comparative Planetology, 8–9 October 2001, Moscow, Russia
- Holme R (1998) Electromagnetic core–mantle coupling. I. Explaining decadal changes in the length of day. *Geophysical Journal International* 132: 167–180.
- Hough SS (1895) The oscillations of a rotating ellipsoidal shell containing fluid. *Philosophical Transactions of the Royal Society of London A* 186: 469–506.
- James PB, Kieffer HH, and Paige DA (1992) The seasonal cycle of carbon dioxide on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 934–968. Tucson, AZ: University of Arizona Press.
- Jault D, Gire D, and Le Mouel JL (1988) Westward drift, core motions and exchanges of angular momentum between core and mantle. *Nature* 333: 353–356.
- Jeffreys H (1952) *The Earth. Its Origin, History, and Physical Constitution*. Cambridge: Cambridge University Press.
- Jehn R, Corral C, and Giampieri G (2004) Estimating Mercury's 88-day libration amplitude from orbit. *Planetary and Space Science* 52: 727–732.
- Karatekin Ö, Van Hoolst T, and Dehant V (2006b) Martian global scale CO₂ exchange from time-variable gravity measurements. *Journal of Geophysical Research* 111: E06003.
- Karatekin Ö, Van Hoolst T, Tastet J, de Viron O, and Dehant V (2006a) The effects of seasonal mass redistribution and interior structure on length-of-day variations of Mars. *Advances in Space Research* 38: 739–744.
- Kaula WM (2000) *Theory of Satellite Geodesy*. New York: Dover.
- Kiefer WS (2003) Melting in the Martian mantle: Shergottite formation and implications for present-day mantle convection on Mars. *Meteoritics and Planetary Science* 38: 1815–1832.
- Kieffer WS, Bills B, and Nerem RS (1996) An inversion of gravity and topography for mantle and crustal structure on Mars. *Journal of Geophysical Research* 101: 9239–9252.
- Kieffer HH, Jakosky BM, and Snyder CW (1992) The planet Mars: From antiquity to the present. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1–33. Tucson, AZ: University of Arizona Press.
- Kinoshita H (1977) Theory of the rotation of the rigid Earth. *Celestial Mechanics* 15: 277–326.
- Klaasen KP (1976) Mercury's rotation axis and period. *Icarus* 28: 469–478.
- Konopliv AS, Banerdt WB, and Sjogren WL (1999) Venus gravity: 180th degree and order model. *Icarus* 139: 3–18.
- Konopliv AS, Borders NJ, Chodas PW, et al. (1993) Venus gravity and topography: 60th degree and order model. *Geophysical Research Letters* 20: 2403–2406.
- Konopliv AS and Sjogren WL (1994) Venus spherical harmonic gravity model to degree and order 60. *Icarus* 112: 42–54.
- Konopliv AS and Yoder CF (1996) Venus k_2 tidal love number from Magellan and PVO tracking data. *Geophysical Research Letters* 23: 1857–1860.
- Konopliv AS, Yoder CF, Standish EM, Yuan D-N, and Sjogren WL (2006) A global solution for the Mars static and seasonal gravity, Mars orientation, Phobos and Deimos masses, and Mars ephemeris. *Icarus* 182: 23–50.
- Kuang W and Bloxham J (1997) On the dynamics of topographical core–mantle coupling. *Physics of the Earth and Planetary Interiors* 99: 289–294.
- Kundt W (1977) Spin and atmospheric tides of Venus. *Astronomy and Astrophysics* 60: 85–91.

- Lago B and Cazenave A (1979) Possible dynamical evolution of the rotation of Venus since formation. *Moon and the Planets* 21: 127–154.
- Lambeck K (1980) *The Earth's Variable Rotation: Geophysical Causes and Consequences*. Cambridge: Cambridge University Press.
- Laskar J (1994) Large scale chaos in the solar system. *Astronomy and Astrophysics* 287: L9–L12.
- Laskar J, Correia ACM, Gastineau M, Joutel F, Levrard B, and Robutel P (2004) Long term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus* 170: 343–364.
- Laskar J and Robutel P (1993) The chaotic obliquity of the planets. *Nature* 361: 608–612.
- Levrard B and Laskar J (2003) Climate friction and the Earth's obliquity. *Geophysical Journal International* 154: 970–990.
- Lewis JS (1988) Origin and composition of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 651–666. Tucson, AZ: The University of Arizona Press.
- Lewis SR, Collins M, Read PL, et al. (1999) A climate database for Mars. *Journal of Geophysical Research* 104(E10): 24177–24194.
- Lissauer JJ, Berman AF, Greenzweig Y, and Kary DM (1997) Accretion of mass and spin angular momentum by a planet on an eccentric orbit. *Icarus* 127: 65–92.
- Lissauer JJ, Dones L, and Ohtsuki K (2000) Origin and evolution of terrestrial planet rotation. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 101–112. Tucson, AZ: University of Arizona Press.
- Lissauer JJ and Kary DM (1991) The origin of the systematic component of planetary rotation. I. Planet on a circular orbit. *Icarus* 93: 126–159.
- Litvak ML, Mitrofanov IG, Kozyrev AS, et al. (2004) Seasonal carbon dioxide depositions on the Martian surface as revealed from neutron measurements by the HEND instrument onboard the 2001 Mars odyssey spacecraft. *Solar System Research* 38: 167–177.
- Litvak ML, Nitrofanov IG, and Sanin AB, et al. (2006) Comparison between polar regions of Mars from HEND/Odyssey data. *Icarus* 180: 23–37 (doi:10.1016/j.icarus.2005.08.009).
- lodders K and Fegley B, Jr. (1998) *The Planetary Scientist's Companion*. New York: Oxford University Press.
- Longhi J, Knittle E, Holloway JR, and Wänke H (1992) The bulk composition, mineralogy and internal structure of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 184–208. Tucson, AZ: University of Arizona Press.
- MacDonald GJF (1964) Tidal friction. *Reviews of Geophysics* 2: 467–541.
- McNamee JB, Borderies NJ, and Sjogren WL (1993) Venus: Global gravity and topography. *Journal of Geophysical Research* 98(E5): 9113–9128.
- Margot, Peale S, Jurgens RF, Slade MA, and Holin IV (2002) Planetary interior structure revealed by spin dynamics. American Geophysical Union, Fall meeting 2002, abstract P22D-08.
- Margot J, Peale S, Jurgens RF, Slade MA, and Holin IV (2004) Earth-based measurements of planetary rotational states. American Geophysical Union, Fall meeting 2004, abstract G33A-02.
- Mathews PM, Buffett BA, Herring TA, and Shapiro II (1991) Forced nutations of the Earth: Influence of inner core dynamics 1. Theory. *Journal of Geophysical Research* 96(B5): 8219–8242.
- Mathews PM and Guo JY (2005) Viscoelectromagnetic coupling in precession-nutation theory. *Journal of Geophysical Research* 110: B02402 (doi:10.1029/2003JB002915).
- Mathews PM, Herring TA, and Buffett BA (2002) Modeling of nutation and precession: New nutation series for nonrigid Earth and insights into the Earth's interior. *Journal of Geophysical Research* 107: ETG 3–1 (doi:10.1029/2001JB000390).
- McCue J and Dormand JR (1993) Evolution of the spin of Venus. *Earth, Moon and Planets* 63: 209–225.
- Melchior P (1983) *The Tides of the Planet Earth*. Oxford, UK: Pergamon.
- Milani A, Rossi A, Vokrouhlický D, Villani D, and Bonanno C (2001) Gravity field and rotation state of Mercury from the BepiColombo radio science experiments. *Planetary and Space Science* 49: 1579–1596.
- Moritz H and Mueller, II (1987) *Earth Rotation. Theory and Observation*. New York: Ungar.
- Mottigner NA, Sjogren WL, and Bills BG (1985) Venus gravity: A harmonic analysis and geophysical implications. *Journal of Geophysical Research* 90: C739–C756.
- Munk WH and MacDonald GJF (1960) *The Rotation of the Earth: A Geophysical Discussion*. Cambridge: Cambridge University Press.
- Murray CD and Dermott SF (1999) *Solar System Dynamics*. Cambridge: Cambridge University Press.
- Nerem RS, Bills BG, and McNamee JB (1993) A high resolution gravity model for Venus: GVM-1. *Geophysical Research Letters* 20: 599–602.
- Néron de Surgy O and Laskar J (1997) On the long term evolution of the spin of the Earth. *Astronomy and Astrophysics* 318: 975–989.
- Ohtsuki K and Ida S (1998) Planetary rotation by accretion of planetesimals with uniform spatial distribution formed by the planet's gravitational perturbation. *Icarus* 131: 393–420.
- Owen T, Biemann K, Rushneck DR, Biller JE, Howarth DW, and Lafleur AL (1977) The composition of the atmosphere at the surface of Mars. *Journal of Geophysical Research* 82: 4635–4639.
- Pais MA, Le Mouél JL, Lambeck K, and Poirier JP (1999) Late Cambrian paradoxical glaciation and obliquity of the Earth – A discussion of dynamical constraints. *Earth and Planetary Science Letters* 174: 155–171.
- Peale SJ (1969) Generalized Cassini laws. *The Astronomical Journal* 74: 483–489.
- Peale SJ (1972) Determination of parameters related to the interior of Mercury. *Icarus* 17: 168–173.
- Peale SJ (1974) Possible histories of the obliquity of Mercury. *The Astronomical Journal* 79: 722–744.
- Peale SJ (1976a) Inferences from the dynamical history of Mercury's rotation. *Icarus* 28: 459–467.
- Peale SJ (1976b) Does Mercury have a molten core? *Nature* 262: 765–766.
- Peale SJ (1988) The rotational dynamics of Mercury and the state of its core. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 461–493. Tucson, AZ: University of Arizona Press.
- Peale SJ (2005) The free precession and libration of Mercury. *Icarus* 178: 4–18.
- Peale SJ (2006) The proximity of Mercury's spin to Cassini state 1 from adiabatic invariance. *Icarus* 181: 338–347.
- Peale SJ and Boss AP (1977) A spin-orbit constraint on the viscosity of a Mercurian liquid core. *Journal of Geophysical Research* 82: 743–749.
- Peale SJ, Phillips RJ, Solomon SC, Smith DE, and Zuber MT (2002) A procedure for determining the nature of Mercury's core. *Meteoritics and Planetary Science* 37: 1269–1283.
- Pettengill GH and Dyce RB (1965) A radar determination of the rotation of the planet Mercury. *Nature* 206: 1240.
- Philip JR (1979) Angular momentum of seasonally condensing atmospheres, with reference to Mars. *Geophysical Research Letters* 6: 727–730.
- Poincaré H (1910) Sur la précession des corps déformables. *Bulletin Astronomique* 27: 321–356.

- Poirier JP (1994) Light elements in the Earth's outer core: A critical review. *Physics of the Earth and Planetary Interiors* 85: 319–337.
- Pollack JB, Haberle RM, Schaeffer J, and Lee H (1990) Simulations of the general circulation of the Martian atmosphere. I: Polar processes. *Journal of Geophysical Research* 95: 1447–1474.
- Rambaux N and Bois E (2004) Theory of the Mercury's spin-orbit motion and analysis of its main librations. *Astronomy and Astrophysics* 413: 381–393.
- Rambaux N, Van Hoolst T, Dehant V, and Bois E (2007) Core-mantle inertial coupling and libration of Mercury. *Astronomy and Astrophysics*, (in press).
- Reasenberg RD and King RW (1979) The rotation of Mars. *Journal of Geophysical Research* 84(B11): 6231–6240.
- Reasenberg RD, Eubanks TM, MacNeil PE, and Shapiro II (1980) Mars rotation: Detection of seasonal rate variation. *Bulletin of the American Astronomical Society* 12: 721.
- Reese CC and Peterson PE (2002) Thermal evolution of Mercury in the conductive regime and implications for magnetism. *Lunar and Planetary Science XXXIII*.
- Roberts JH and Zhong SJ (2004) Plume-induced topography and geoid anomalies and their implications for the Tharsis rise on Mars. *Journal of Geophysical Research* 109: E03009.
- Roberts PH and Stewartson K (1965) On the motion of a liquid in a spheroidal cavity of a precessing rigid body. II. *Proceedings of the Cambridge Philosophical Society* 61: 279–288.
- Rochester MG (1976) The secular decrease of obliquity due to dissipative core–mantle coupling. *Geophysical Journal of the Royal Astronomical Society* 46: 109–126.
- Roosbeek F (1999) Analytical developments of rigid Mars nutation and tide generating potential series. *Celestial Mechanics and Dynamical Astronomy* 75: 285–300.
- Rubincam DP (1990) Mars: Change in axial tilt due to the climate? *Science* 248: 720–721.
- Russell CT, Baker DN, and Slavin JA (1988) The magnetosphere of Mercury. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 514–561. Tucson, AZ: University of Arizona Press.
- Sanchez B, Haberle R, and Schaeffer J (2004) Atmospheric rotational effects on Mars based on the NASA Ames general circulation model: Angular momentum approach. *Journal of Geophysical Research* 109: E08005 (doi:10.1029/2004JE002254).
- Sanchez BV, Rowlands DD, Haberele RM, and Schaeffer J (2003) Atmospheric rotational effects on Mars based on the NASA Ames general circulation model. *Journal of Geophysical Research* 108(E5): 5040 (doi:10.1029/2002JE001984).
- Sasao T, Okamoto I, and Sakai S (1977) Dissipative core–mantle coupling and nutational motion of the Earth. *Publications of the Astronomical Society of Japan* 29: 83–105.
- Sasao T, Okubo S, and Saito M (1980) A simple theory on the dynamical effects of a stratified fluid core upon nutational motion of the Earth. In: Fedorov EP, Smith ML, and Bender PL (eds.) *Proceedings of IQU Symposium 78: Nutation and the Earth's Rotation*, Kiev, May 1977, pp. 165–183. Dordrecht: Reidel.
- Schubert G, Ross MN, Stevenson DJ, and Spohn T (1988) Mercury's thermal history and the generation of its magnetic field. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 429–460. Tucson, AZ: University Arizona Press.
- Schubert G, Solomon SC, Turcotte DL, Drake MJ, and Sleep NH (1992) Origin and thermal evolution of Mars. In: Kiebler HH, Jakosky BM, Snyder CW, and Mathews MS (eds.) *Mars*, pp. 147–183. Tucson, AZ: University of Arizona Press.
- Schubert G and Spohn T (1990) Thermal history of Mars and the sulfur content of its core. *Journal of Geophysical Research* 95(B9): 14095–14104.
- Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*. Cambridge: Cambridge University Press.
- Schumacher S and Breuer D (2005) An alternative explanation for recent volcanism on Mars. American Geophysical Union, Fall meeting 2005, abstract #P51B–0926.
- Shapiro II (1967) Resonance rotation of Venus. *Science* 157: 423–425.
- Shapiro, II, Chandler JF, Campbell DB, Hine AA, and Stacy NJS (1990) The spin vector of Venus. *The Astronomical Journal* 100: 1363–1368.
- Shen M and Zhang CZ (1989) Dynamical evolution of the rotation of Venus. *Earth, Moon and Planets* 43: 275–287.
- Sjogren WL, Banerdt BA, Chodas PW, et al. (1997) The Venus gravity field and other geodetic parameters. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere and Solar Wind Environment*, pp. 1125–1161. Tucson, AZ: University of Arizona Press.
- Smaan WM (1953) *Celestial Mechanics*. London: Longmans, Green.
- Smith DE, Zuber MT, Haberle RM, Rowlands DD, and Murphy JR (1999) The Mars seasonal CO₂ cycle and the time variation of the gravity field: A general circulation model simulation. *Journal of Geophysical Research* 104: 1885–1896.
- Smith DE, Zuber MT, and Neumann GA (2001) Seasonal variations of snow depth on Mars. *Science* 294: 2141–2146.
- Smith JC and Born GH (1976) Secular acceleration of Phobos and Q of Mars. *Icarus* 27: 51–53.
- Smith ML (1977) Wobble and nutation of the Earth. *Geophysical Journal of the Royal Astronomical Society* 50: 103–140.
- Smith ML and Dahlen FA (1981) The period and Q of the Chandler wobble. *The Geophysical Journal* 64: 223–281.
- Sohl F and Spohn T (1997) The structure of Mars: Implications from SNC-meteorites. *Journal of Geophysical Research* 102: 1613–1635.
- Solomon SC, McNutt RL, Jr., Gold RE, et al. (2001) The MESSENGER mission to Mercury: Scientific objectives and implementation. *Planetary and Space Science* 49: 1445–1465.
- Spada G and Alfonsi L (1998) Obliquity variations due to climate friction on Mars: Darwin versus layered models. *Journal of Geophysical Research* 103(E12): 28599–28605.
- Spohn T (1991) Mantle differentiation and thermal evolution of Mars, Mercury, and Venus. *Icarus* 90: 222–237.
- Spohn T, Sohl F, Wieczkowski K, and Conzelmann V (2001) The interior structure of Mercury: What we know, what we expect from BepiColombo. *Planetary and Space Science* 49: 1561–1570.
- Stacey FD (1992) *Physics of the Earth*. Brisbane, Australia: Brook-Field Press.
- Stanley S, Bloxham J, Hutchinson WE, and Zuber MT (2005) Thin shell dynamo models consistent with Mercury's weak observed magnetic field. *Earth and Planetary Science Letters* 234: 27–38.
- Stevenson DJ (2001) Mars core and magnetism. *Nature* 421: 214–219.
- Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.
- Stewartson K and Roberts PH (1963) On the motion of a liquid in a spheroidal cavity of a precessing rigid body. *Journal of Fluid Mechanics* 17: 1–20.
- Szeto AMK and Xu S (1997) Gravitational coupling in a triaxial ellipsoidal Earth. *Journal of Geophysical Research* 102: 27651–27657.
- Thompson W (1882) On the thermodynamic acceleration of the Earth's rotation. *Proceedings of the Royal Society of London Edinburg* 11: 396–405.

- Thomson DJ (1990) Quadratic inverse spectrum estimates: Applications to paleoclimatology. *Philosophical Transactions of the Royal Society of London A* 332: 539–597.
- Touma J and Wisdom J (1993) The chaotic obliquity of Mars. *Science* 259: 1294–1297.
- Van den Acker E, Van Hoolst T, de Viron O, et al. (2002) Influence of the winds and the CO₂ mass exchange between the atmosphere and the polar ice caps on Mars' orientation parameters. *Journal of Geophysical Research* doi:10.1029/2000JE001539.
- Van Hoolst T and Dehant V (2002) Influence of triaxiality and second-order terms in flattening on the rotation of terrestrial planets. I: Formalism and rotational normal modes. *Physics of the Earth and Planetary Interiors* 134: 17–33.
- Van Hoolst T, Dehant V, and Defraigne P (2000a) Sensitivity of the free core nutation and the Chandler wobble to changes in the interior structure of Mars. *Physics of the Earth and Planetary Interiors* 117: 397–405.
- Van Hoolst T, Dehant V, and Degraigne P (2000b) Chandler wobble and free core nutation for Mars. *Planetary and Space Science* 48: 1145–1151.
- Van Hoolst T, Dehant V, Roosbeek F, and Lognonn P (2003) Tidally induced surface displacements, external potential variations, and gravity variations on Mars. *Icarus* 1612: 281–296.
- Van Hoolst T and Jacobs C (2003) Mercury's tides and interior structure. *Journal of Geophysical Research* 108(E11): 5121 (doi:10.1029/2003JE002126).
- Wahr JM (1981a) A normal mode expansion for the forced response of a rotating Earth. *Geophysical Journal of the Royal Astronomical Society* 64: 651–675.
- Wahr JM (1981b) The forced nutations of an elliptical, rotating, elastic and oceanless Earth. *Geophysical Journal of the Royal Astronomical Society* 64: 705–727.
- Ward WR (1974) Climatic variations on Mars. 1. Astronomical theory of insulation. *Journal of Geophysical Research* 79: 3375–3386.
- Ward WR (1975) Tidal friction and generalized Cassini's laws in the solar system. *The Astronomical Journal* 80: 64–70.
- Ward WR (1992) Long-term orbital dynamics and spin dynamics of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 298–320. Tucson, AZ: University of Arizona Press.
- Wetherill GW (1988) Accumulation of Mercury from planetesimals. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 670–691. Tucson AZ: University of Arizona Press.
- Williams DM, Kasting JF, and Frakes LA (1998) Low-latitude glaciation and rapid changes in the Earth's obliquity explained by obliquity–oblateness feedback. *Nature* 396: 453–455.
- Yoder CF (1982) Tidal rigidity of Phobos. *Icarus* 49: 327–346.
- Yoder CF (1995) Venus' free obliquity. *Icarus* 117: 250–286.
- Yoder CF (1997) Venusian spin dynamics. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II. Geology, Geophysics, Atmosphere and Solar Wind Environment*. Tucson, AZ: University of Arizona Press.
- Yoder CF, Konopliv AS, Yuan DN, Standish EM, and Folkner WM (2003) Fluid core size of Mars from detection of the solar tide. *Science* 300: 299–303 (doi:10.1126/science.1079645).
- Yoder CF and Standish EM (1997) Martian precession and rotation from Viking lander range data. *Journal of Geophysical Research* 102(E2): 4065–4080.
- Yoder CF and Ward WR (1979) Does Venus wobble? *The Astrophysical Journal* 233: L33–L37.
- Yoder CF, Williams JG, and Parke ME (1981) Tidal variations of Earth rotation. *Journal of Geophysical Research* 86: 881–891.
- Yseboodt M, Barriot J-P, and Dehant V (2003) Analytical modeling of the Doppler tracking between a lander and a Mars orbiter in terms of rotational dynamics. *Journal of Geophysical Research* 108(E7): 5075 (doi:10.1029/2003JE002045).
- Yseboodt M and Margot J-L (2006) Evolution of Mercury's obliquity. *Icarus* 181: 327–337.
- Zharkov VN and Gudkova TV (2005) Construction of martian interior model. *Solar System Research* 39: 343–373.

10.05 Gravity and Topography of the Terrestrial Planets

M. A. Wieczorek, Institut de Physique du Globe de Paris, Saint Maur des Fossés, France

© 2007 Elsevier B.V. All rights reserved.

10.05.1	Introduction	165
10.05.2	Mathematical Preliminaries	166
10.05.2.1	Spherical Harmonics	167
10.05.2.2	The Potential, Gravity, and Geoid	168
10.05.3	The Data	171
10.05.3.1	Earth	171
10.05.3.1.1	Topography	171
10.05.3.1.2	Gravity	173
10.05.3.1.3	Spectral analysis	174
10.05.3.2	Venus	174
10.05.3.2.1	Topography	174
10.05.3.2.2	Gravity	175
10.05.3.2.3	Spectral analysis	177
10.05.3.3	Mars	178
10.05.3.3.1	Topography	178
10.05.3.3.2	Gravity	178
10.05.3.3.3	Spectral analysis	180
10.05.3.4	The Moon	181
10.05.3.4.1	Topography	181
10.05.3.4.2	Gravity	183
10.05.3.4.3	Spectral analysis	183
10.05.4	Methods for Calculating Gravity from Topography	184
10.05.5	Crustal Thickness Modeling	186
10.05.6	Admittance Modeling	187
10.05.6.1	Spatial Domain	187
10.05.6.2	Spectral Domain	189
10.05.7	Localized Spectral Analysis	192
10.05.8	Summary of Major Results	193
10.05.8.1	Earth	193
10.05.8.2	Venus	195
10.05.8.3	Mars	197
10.05.8.4	The Moon	198
10.05.9	Future Developments and Concluding Remarks	199
	References	201

10.05.1 Introduction

One of the most precise ways to investigate the subsurface structure of a planet is through the analysis of seismic waves. While such endeavors have proved to be remarkably successful for the Earth, the emplacement of a geophysical network that includes seismometers on a terrestrial body such as Mercury, Venus, Mars, or the Moon is both costly and technologically challenging (*see* Chapters 10.02,

10.03, and 10.18). Fortunately, alternative means exist to probe the interior structure of the terrestrial planets from orbit. One such method, reviewed in Chapter 10.04, is through analyses of a planet's time variable rotation. Such investigations can measure the moment of inertia factor (when combined with low-degree gravity measurements) and k_2 Love number, from which constraints on the density and viscosity profile can be obtained. The magnetic induction response of a planet to time variable

magnetic fields can be used to determine how the body's electrical conductivity (and hence composition) varies with depth. Finally, because the gravitational field of a planet is sensitive to its internal density structure, another powerful method that can be used to probe the subsurface, and which is the subject of this review, is the joint analysis of gravitational and topographic data.

The past decade has witnessed dramatic advancements in our understanding and knowledge of the gravitational fields and topography of the terrestrial planets. These advances are intimately related to the acquisition of radio tracking data from orbiting spacecraft (which can be used to invert for the gravitational field), as well as the collection of data from orbiting altimeters. As examples, the first near-global topographic map of the Moon was obtained by the Clementine mission in 1994 ([Zuber et al., 1994](#)), and the topography of Venus and Mars was mapped to high resolution by the Magellan ([Ford and Pettengill, 1992](#)) and Mars Global Surveyor (MGS) ([Smith et al., 1999](#)) missions which were placed in orbit in 1990 and 1997, respectively. In addition, the topography and gravitational field of the Earth has been continually refined by a series of recent and ongoing missions.

From a geophysical perspective, the motivation for obtaining high-resolution gravitational and topographic data is to place constraints on the interior structure of a planet. As a result of Newton's law of gravitation, the magnitude and direction of the gravitational acceleration exterior to a planet is completely determined by its internal distribution of mass. When combined with topographic data and reasonable geologic assumptions, it is possible to invert for important geophysical parameters such as crustal thickness, elastic thickness, and crustal and mantle density. These model parameters can then be used to address questions concerning planetary differentiation, crust formation, thermal evolution, and magmatic processes. As the spatial resolving power of gravitational measurements decreases with increasing distance from the object, such investigations are generally restricted to the crust and upper mantle.

Very few research articles have been written that review the gravitational fields and topography of the terrestrial planets from a comparative planetology perspective: exceptions include [Phillips and Lambeck \(1980\)](#), [Balmino \(1993\)](#), [Bills and Lemoine \(1995\)](#), and [Rummel \(2005\)](#). The purpose of this chapter is to both review the current state of knowledge of the gravitational fields and topography of the terrestrial planets, and to review the tools that are

used to describe and analyze these data. While gravitational and topographic data sets can each be used independently to make inferences about the interior structure of a planet, such results are often based upon untestable hypotheses and/or are highly under-constrained. Thus, although regional topographic models have been constructed for some planets, moons, and asteroids (such as Mercury (e.g., [Harmon et al., 1986](#); [Anderson et al., 1996a](#); [Watters et al., 2001, 2002](#); [André et al., 2005](#)), Ganymede ([Giese et al., 1998](#)), Europa ([Nimmo et al., 2003a, 2003b](#)), Phobos ([Duxbury, 1989](#)), 433 Eros ([Zuber et al., 2000b](#)), and Itokawa ([Abe et al., 2006](#))), and the longest wavelength gravitational fields have been constrained for others (such as Mercury ([Anderson et al., 1987](#)), Eros ([Miller et al., 2002](#); [Garmier and Barriot, 2002](#)), Io ([Anderson et al., 1996c, 2001a](#)), Europa ([Anderson et al., 1998](#)), Ganymede ([Anderson et al., 1996b](#)), and Callisto ([Anderson et al., 2001b](#))), this chapter focuses on those bodies for which the gravity and topography are both characterized to high degree, namely, the Earth, Venus, Mars, and the Moon.

The outline of this chapter is as follows. First, in Section 10.05.2, a general review of the mathematical formalism that is used in describing the properties of gravitational fields and topography is given. Next, in Section 10.05.3, the basic properties of the gravitational fields and topography of the Earth, Venus, Mars, and the Moon are characterized. Following this, Sections 10.05.4–10.05.7 discuss the relationship between gravity and topography, and how the two data sets can be used to invert for geophysical parameters. These methods include crustal thickness modeling, the analysis of geoid/topography ratios (GTRs), and modeling of the spectral admittance and correlation functions. Section 10.05.8 summarizes the major results that have been obtained by gravity and topography modeling for these planetary bodies, and Section 10.05.9 concludes by discussing future developments that can be expected in this domain.

10.05.2 Mathematical Preliminaries

This section reviews the mathematical background that is required to understand how gravitational fields and topography are characterized and manipulated. As most analysis techniques make use of spherical harmonics, the first subsection defines these functions and introduces certain concepts such as the power spectrum. The following subsection

gives definitions that are specific to the gravitational potential, gravitational field, and geoid. For further details, the reader is referred to several key books and articles such as Jeffreys (1976), Kaula (1967, 2000), Heiskanen and Moritz (1967), and Lambeck (1988).

10.05.2.1 Spherical Harmonics

Spherical harmonics are the natural set of orthogonal basis functions on the surface of the sphere. As such, any real square-integrable function can be expressed as a linear combination of these as

$$f(\Omega) = \sum_{l=0}^{\infty} \sum_{m=-l}^{l} f_{lm} Y_{lm}(\Omega) \quad [1]$$

where Y_{lm} is the spherical harmonic function of degree l and order m , f_{lm} is the corresponding expansion coefficient, and $\Omega = (\theta, \phi)$ represents position on the sphere in terms of colatitude θ and longitude ϕ . Alternative representations include the use of ellipsoidal harmonics (e.g., Garmier and Barriot, 2001) when the function is globally known, and spherical cap harmonics (e.g., Haines, 1985; Hwang and Chen, 1997; Thébault *et al.*, 2004, 2006) when the function is known only regionally. In geodesic and gravitational applications, the real spherical harmonics are defined as

$$Y_{lm}(\Omega) = \begin{cases} \bar{P}_{lm}(\cos \theta) \cos m\phi & \text{if } m \geq 0 \\ \bar{P}_{l|m|}(\cos \theta) \sin |m|\phi & \text{if } m < 0 \end{cases} \quad [2]$$

where the normalized associated Legendre functions are given by

$$\bar{P}_{lm}(\mu) = \sqrt{(2 - \delta_{0m})(2l + 1) \frac{(l-m)!}{(l+m)!}} P_{lm}(\mu) \quad [3]$$

and where δ_{ij} is the Kronecker delta function. The unnormalized Legendre functions in the above equation are defined in relation to the Legendre polynomials by

$$P_{lm}(\mu) = (1 - \mu^2)^{m/2} \frac{d^m}{d\mu^m} P_l(\mu) \quad [4]$$

$$P_l(\mu) = \frac{1}{2^l l!} \frac{d^l}{d\mu^l} (\mu^2 - 1)^l \quad [5]$$

The normalized associated Legendre functions are orthogonal for a given value of m according to

$$\int_{-1}^1 \bar{P}_{lm}(\mu) \bar{P}_{l'm'}(\mu) d\mu = 2(2 - \delta_{0m}) \delta_{ll'} \quad [6]$$

and the spherical harmonics are orthogonal over both l and m with the normalization

$$\int_{\Omega} Y_{lm}(\Omega) Y_{l'm'}(\Omega) d\Omega = 4\pi \delta_{ll'} \delta_{mm'} \quad [7]$$

where $d\Omega = \sin \theta d\theta d\phi$. Multiplying eqn [1] by $Y_{l'm'}$ and integrating over all space, the spherical harmonic coefficients of the function f can be obtained by calculating the integral

$$f_{lm} = \frac{1}{4\pi} \int_{\Omega} f(\Omega) Y_{lm}(\Omega) d\Omega \quad [8]$$

A useful visualization property of the spherical harmonic functions is that they possess $2|m|$ zero crossings in the longitudinal direction, and $l - |m|$ zero crossings in the latitudinal direction. In addition, for a given spherical harmonic degree l , the equivalent Cartesian wavelength is $\lambda \approx 2\pi R / \sqrt{l(l+1)}$, where R is the mean planetary radius, a result known as the Jeans relation. It should be noted that those coefficients and spherical harmonics where $m = 0$ are referred to as zonal, those with $l = |m|$ are sectoral, and the rest are tesseral. Furthermore, the zonal coefficients C_{l0} are often written as $-\mathcal{J}_l$.

Using the orthogonality properties of the spherical harmonic functions, it is straightforward to verify that the total power of a function f is related to its spectral coefficients by a generalization of Parseval's theorem:

$$\frac{1}{4\pi} \int_{\Omega} f^2(\Omega) d\Omega = \sum_{l=0}^{\infty} S_{ff}(l) \quad [9]$$

where

$$S_{ff}(l) = \sum_{m=-l}^{l} f_{lm}^2 \quad [10]$$

is referred to as the power spectrum of the function. (The term ‘power’ is here used in the sense of the signal processing literature, namely, the square of the function divided by the area that it spans.) It can be shown that S_{ff} is unmodified by a rotation of the coordinate system. Similarly, the cross-power of two functions f and g is given by

$$\frac{1}{4\pi} \int_{\Omega} f(\Omega) g(\Omega) d\Omega = \sum_{l=0}^{\infty} S_{fg}(l) \quad [11]$$

where

$$S_{fg}(l) = \sum_{m=-l}^{l} f_{lm} g_{lm} \quad [12]$$

is defined as the cross-power spectrum. If the functions f and g have a zero mean (i.e., their zero-degree terms are equal to zero), then $S_{ff}(l)$ and $S_{fg}(l)$ represent the contributions to the variance and covariance, respectively, for degree l . Some authors plot routinely the power per degree, $S/(2l+1)$, or the rms amplitude of the spherical harmonic coefficients, $\sqrt{S/(2l+1)}$.

One source of confusion with spherical harmonic analyses is that not all authors use the same definitions for the spherical harmonic and Legendre functions. In contrast to the ‘geodesy’ or ‘ 4π ’ normalization of eqn [7] (cf. Kaula, 2000), the seismology (e.g., Dahlen and Tromp, 1998) and physics (e.g., Varshalovich *et al.*, 1988) communities often use orthonormal harmonics, whose squared integral is equal to unity. Alternatively, the geomagnetic community employs Schmidt semi-normalized harmonics whose squared integral is $4\pi/(2l+1)$ (e.g., Blakely, 1995). A more subtle problem is related to a phase factor of $(-1)^m$ (the Condon–Shortley phase) that is sometimes applied to either eqns [3] or [4]. While the spherical harmonics used by the geodesy and geomagnetic communities both exclude this phase factor, those in the physics and seismology communities often include it.

In order to obtain the spherical harmonic coefficients f_{lm} of a function f , it is necessary to be able to calculate accurately both the normalized Legendre functions of eqn [3] and the integral of eqn [8]. Methods for efficient calculation of the normalized associated Legendre functions depend upon the use of well-known three-term recursion formulas. If starting values used in the recursion are appropriately scaled, as is summarized by Holmes and

Featherstone (2002), these can be computed to high accuracy up to a maximum spherical harmonic degree of about 2700. To obtain a similar accuracy at higher degrees would require the use of an alternative scaling algorithm.

The integrals of eqn [8] are most easily performed if the function f is known on a set of evenly spaced grid points in longitude. Numerical methods for calculating this integral generally involve Fourier-transforming each latitudinal band, and then integrating over latitude for each l and m (e.g., Sneeuw, 1994). If the function is sampled on an $n \times n$ grid, with n even, and if the function is known to be band-limited to a maximum degree $n/2 - 1$, then the spherical harmonic transform can be computed exactly (see Driscoll and Healy, 1994). Alternatively, the integral over latitude can be performed using Gauss–Legendre quadrature (e.g., Press *et al.*, 1992, pp. 140–149). While the integrand of eqn [8] is not in general a terminating polynomial, if the function is band-limited to a maximum degree n , $n + 1$ points in latitude are sufficient to calculate accurately the spherical harmonic expansion coefficients. Software for performing spherical harmonic transforms and reconstructions is available at the author’s website (see Table 1 for a list of internet resources).

10.05.2.2 The Potential, Gravity, and Geoid

If the gravitational acceleration \mathbf{g} is expressed as the gradient of a scalar potential U ,

$$\mathbf{g}(\mathbf{r}) = \nabla U(\mathbf{r}) \quad [13]$$

Table 1 Internet resources

Resource	Internet address
SHTOOLS: Fortran 95 spherical harmonics code	http://www.ipgp.jussieu.fr/~wieczor/
Planetary Data System (PDS) Geosciences node	http://pds-geosciences.wustl.edu/
ETOPO2: Earth topography model	http://www.ngdc.noaa.gov/mgg/fliers/01mgg04.html
GTOPO30: Earth topography model	http://edcdaac.usgs.gov/gtopo30/gtopo30.asp
SRTM: Earth topography model	http://srtm.usgs.gov/ and ftp://e0srp01u.ecs.nasa.gov
STRM30_PLUS: Earth topography model	http://topex.ucsd.edu/WWW_html/srtm30_plus.html
WGS84 ellipsoid and ‘WGS84 EGM96 geoid’	http://earth-info.nga.mil/GandG/wgs84/
GGM02: Earth gravity model	http://www.csr.utexas.edu/grace/gravity/
EIGEN: Earth gravity models	http://www.gfz-potsdam.de/pb1/op/grace/results/
GTDR3.2: Venus topography	http://voir.mit.edu/pub/mg_3003/
Planetary maps with feature names	http://ralphaeschliman.com/
Generic Mapping Tools	http://gmt.soest.hawaii.edu/

Newton's law can be used to calculate the gravitational potential at an arbitrary point by a simple integral over the mass distribution,

$$U(\mathbf{r}) = \int_V \frac{G\rho(\mathbf{r}')}{|\mathbf{r}-\mathbf{r}'|} dV' \quad [14]$$

where \mathbf{r} represents position, ρ is the mass density, and V signifies that space which is occupied by the body. While the sign convention of the above two equations is consistent with that used in the geodesy literature, it should be noted that other disciplines, such as physics, place a negative sign in front of each of these. Exterior to the mass distribution V , it can be shown that the potential satisfies Laplace's equation (e.g., Kaula, 2000, ch. 1):

$$\nabla^2 U(\mathbf{r}) = 0 \quad [15]$$

By use of this relationship and the method of separation of variables, the potential U exterior to V can be expressed alternatively as a sum of spherical harmonic functions:

$$U(\mathbf{r}) = \frac{GM}{r} \sum_{l=0}^{\infty} \sum_{m=-l}^l \left(\frac{R_0}{r}\right)^l C_{lm} Y_{lm}(\Omega) \quad [16]$$

Here, the C_{lm} s represent the spherical harmonic coefficients of the gravitational potential at a reference radius R_0 , G is the gravitational constant, and M is the total mass of the object. In practice, the infinite sum is truncated beyond a maximum degree L that is justified by the data resolution. The coefficients C_{lm} of eqn [16] are uniquely related to the internal mass distribution of the body, and methods for calculating these are deferred until Section 10.05.4. Here, it is only noted that the zero-degree coefficient C_{00} is equal to unity, and that if the coordinate system has been chosen such that it coincides with the center of mass of the object, the first-degree terms (C_{10} , C_{11} , and C_{1-1}) are identically zero.

As a result of the factor $(R_0/r)^l$ that multiplies each term in eqn [16] the high-frequency components in this sum (i.e., those with large l) will be attenuated when the observation radius r is greater than the reference radius R_0 . In contrast, if the coefficients were determined at the altitude of an orbiting satellite, and if this equation were used to determine the potential field at the surface of the planet, then the high frequency terms would instead be relatively amplified. Since the spherical harmonic coefficients always possess some uncertainty, which generally increases with increasing l , the process of downward continuing a potential field is not stable and must generally be compensated by some form of filtering

(e.g., Phipps Morgan and Blackman, 1993; Wieczorek and Phillips, 1998; Fedi and Florio, 2002).

If the body under consideration is rotating, an additional non-gravitational force must be taken into account when one is on the surface. In the reference frame of the rotating body, this can be accounted for by adding to eqn [16] a pseudo-potential term that is a result of the centrifugal force. This rotational potential, as well as its spherical harmonic expansion, is given by

$$U^{\text{rot}} = \frac{\omega^2 r^2 \sin^2 \theta}{2} = \omega^2 r^2 \left(\frac{1}{3} Y_{00} - \frac{1}{3\sqrt{5}} Y_{20} \right) \quad [17]$$

where ω is the angular velocity of the rotating object. For some applications, especially concerning the Earth and Moon, it is necessary to include the tidal potential of the satellite or parent body when calculating the potential (see Zharkov *et al.*, 1985; Dermott, 1979). For a synchronously locked satellite on a circular orbit, such as the Moon, the combined tidal and rotational potential in the rotating frame is (ignoring degree-0 terms that do not depend upon r) given approximately by

$$U^{\text{tide + rot}} \approx \omega^2 r^2 \left(\frac{1}{3} Y_{00} - \frac{5}{6\sqrt{5}} Y_{20} + \frac{1}{4} \sqrt{\frac{12}{5}} Y_{22} \right) \quad [18]$$

An important quantity in both geodesy and geophysics is the geoid, which is defined to be a surface that possesses a specific value of the potential. (Although one could come up with imaginative names for equipotential surfaces on Venus, Eros, and Io, among others, the term 'geoid' will here be used for all.) As there are no tangential forces on an equipotential surface (see eqn [13]), a static fluid would naturally conform to the geoid. The oceans of the Earth are approximately static and are well approximated by such a surface. In geophysics, stresses within the lithosphere are often calculated by referencing the vertical position of a density contrast to an equipotential surface. This is necessary when performing lithospheric flexure calculations, especially when considering the longest wavelengths.

The height N of an equipotential surface above a spherical reference radius R can be obtained by approximating the potential by a Taylor series

$$U(R+N) \approx U(R) + \frac{dU(R)}{dr} N + \frac{1}{2} \frac{d^2 U(R)}{dr^2} N^2 \quad [19]$$

and setting this expression equal to a constant

$$U(R+N) = \frac{GM}{R} + \frac{\omega^2 R^2}{3} \quad [20]$$

the value of which is here chosen to be the degree-0 term of the potential at the reference radius R for a rotating planet. Since this equation is quadratic in N , the geoid height can be solved for analytically at any given position. Analytic expressions for the partial derivatives of the potential are easily obtained in the spectral domain from eqns [16], [22] and [17], and spatial renditions of these can be obtained by performing the inverse spherical harmonic transform.

For most cases it is sufficient to use only the first order terms of eqn [19]. Specifically, if the first derivative of U is approximated by $-GM/R^2$, then the geoid is simply given by

$$N \approx R \sum_{l=2}^{\infty} \sum_{m=-l}^l \left(\frac{R_0}{R} \right)^l C_{lm} Y_{lm} - \frac{\omega^2 R^4}{3\sqrt{5}GM} Y_{20} \quad [21]$$

where the degree-1 terms have been assumed to be zero. Although the maximum difference between the first and second order methods is less than 3 m for both the Moon and Venus, differences of up to 20 and 40 m are obtained for the Earth and Mars, respectively. The poor performance of the first-order approximation for the Earth and Mars is primarily a result of the large rotationally induced \mathcal{J}_2 potential terms.

Despite the simplicity of the above method for obtaining the height to an equipotential surface, the question arises as to which equipotential surface should be used. For the Earth, a natural choice is the potential corresponding to mean sea level. However, for the other planets, the choice is more arbitrary. As the above equations for calculating the potential are strictly valid only when the observation point is exterior to the body, one manner of picking a specific potential might be to chose that value for which all points on the geoid are exterior to the body. (Calculation of the potential below the surface of a planet would require knowledge of the interior density distribution.) Another approach might be to use the mean potential on the planet's equator.

The radial component of the gravitational field is easily calculated by taking the first radial derivative of eqn [16]. Ignoring the rotational potential, and using the sign convention that gravitational accelerations are positive when directed downward, this is given by the expression

$$g_r = \frac{GM}{r^2} \sum_{l=0}^{\infty} \sum_{m=-l}^l \left(\frac{R_0}{r} \right)^l (l+1) C_{lm} Y_{lm} \quad [22]$$

Note that this equation differs from that of the potential only by the inclusion of the additional factors $1/r$ and $(l+1)$. The second factor gives a greater importance to the higher degree terms, and it is for this reason that plots of the potential and geoid appear to be 'smooth' when compared to components of the gravitational field. In terrestrial applications it is common to calculate the gravitational field on the geoid. By inserting eqn [21] into eqn [22], and ignoring rotational and higher-order terms, it can be shown that the radial component of the gravitational field on the geoid can be calculated simply by replacing the term $(l+1)$ in eqn [22] by $(l-1)$. The standard unit for quantifying gravitational perturbations is the Galileo, where $1 \text{ Gal} = 10^{-2} \text{ m s}^{-2}$, and when plotting gravitational anomalies in map form, it is conventional to use mGals.

Finally, it will be useful to characterize the relationship between the gravitational field and topography in the spectral domain. For this, let us presume that the radial gravity g_m and topography b_{lm} are related by

$$g_m = Q_{lm} b_{lm} + I_{lm} \quad [23]$$

where Q_{lm} is a linear nonisotropic transfer function, and I_{lm} is that portion of the gravitational signal, such as noise, that is not predicted by the model (the topography can generally be considered noise free). If I_{lm} is uncorrelated with the topography, then it is straightforward to show that the expectations of both S_{bb} and S_{bg} will be unaffected by this signal. (This can be shown by multiplying eqn [23] by b_{lm} , summing over all m , and taking the expectation.) In contrast, the expectation of the gravitational power spectrum will depend upon the magnitude of I_{lm} . In particular, for the case where I_{lm} is a random variable that is independent of the surface topography, the expectation of the gravitational power spectrum is simply

$$S_{gg}(l) = S_{gg}^0(l) + S_{II}(l) \quad [24]$$

where S_{gg}^0 is the power spectrum predicted exclusively by the model Q_{lm} and S_{II} is the power spectrum of I_{lm} . Thus, gravitational measurement noise will bias the gravitational power spectrum by an additive constant.

In quantifying the relationship between gravity and topography it is useful to work with ratios of their cross-power spectra. One quantity, referred to as the admittance, is defined by the cross-power of

the two functions divided by the power of the topography (e.g., Dorman and Lewis, 1970):

$$Z(l) = \frac{S_{bg}(l)}{S_{bb}(l)} \quad [25]$$

This function is not biased by the presence of noise, and when the transfer function Q_{lm} is isotropic (i.e., independent of m), it is an unbiased estimate of Q_l . Another ratio that is commonly used, which is dimensionless, is

$$\gamma(l) = \frac{S_{bg}(l)}{\sqrt{S_{bb}(l)S_{gg}(l)}} \quad [26]$$

If for a given degree l the coefficients b_{lm} and g_{lm} possess a zero mean, the above function is simply the degree correlation of the two sets of harmonic coefficients, which can possess extreme values of 1 and -1 . While noise will not bias the admittance function, it will bias the degree correlation function toward lower values as a consequence of the gravitational power spectrum in the denominator of this expression. The term coherence is usually reserved for the correlation squared, but this definition is not universally followed. Since squaring the correlation discards its sign, its use is not advocated here. It is noted that the isotropic version of eqn [23] predicts that the spectral correlation coefficient should always be ± 1 . Nevertheless, as discussed in Section 10.05.7, nonisotropic models yield expressions that are wavelength-dependent. While other ratios of cross-power spectra could be constructed, only two will be independent.

10.05.3 The Data

10.05.3.1 Earth

10.05.3.1.1 Topography

Despite the fact that the measurement of the Earth's topography and bathymetry has been the subject of numerous government supported campaigns, large portions of the Earth's surface, namely the oceans, remain poorly characterized. Indeed, from a global perspective, it can be said that the topography of Venus and Mars is better known than that of the planet we call home. Until recently, even the elevations of the aerial portions of the continents possessed long-wavelength uncertainties, a result of mosaicking numerous elevation models, each with its own reference surface, along artificial political boundaries. While major advances have been made in the past

decade towards generating global models, the main deficiency is still the sparse bathymetry of the oceans.

Numerous topographic models of the Earth's landmass have been assembled from various sources over the past few decades, including ETOPO5 and ETOPO2 (5- and 2-arcminute resolution, National Geophysical Data Center, 2001), GLOBE (30 arcsec resolution, see Hastings and Dunbar, 1999) and GTOPO30 (detailed documentation for these and all following topography models can be found at the appropriate web address listed in Table 1). Currently, the most accurate model of the Earth's landmass comes from radar interferometric data collected by the Shuttle Radar Topography Mission (SRTM). During its ten days of operation onboard the US space shuttle in year 2000, this mission mapped nearly 80 percent of the landmass between 60° N and 54° S with a horizontal sampling of 1 arcsec (~ 30 m) and an absolute vertical accuracy better than about 10 m (Rabus *et al.*, 2003; Rodríguez *et al.*, 2005). Because of the 5.6 cm wavelength of the radar, elevations generally correspond to the top of the canopy when vegetation is present.

The bathymetry of the oceans has been measured from ship surveys using echo sounding for over half a century. Unfortunately, the ship tracks sometimes possess large navigational errors, and large data gaps exist. As reviewed by Marks and Smith (2006), many data set exist that are based upon these measurements, but each has its own peculiarities. In the absence of additional ship survey data, one method that can be used to improve the bathymetry of the oceans is by combining ship survey data with marine gravity data. As is detailed by Smith and Sandwell (1994, 1997), marine gravity anomalies (as obtained from altimeter derived sea surface slopes) are highly correlated with seafloor topography over a restricted wavelength band. By combining predicted topography from bandpassed gravity anomalies with the long wavelength bathymetry from shipboard sounding, it has been possible to obtain near-global 2-arcsec resolution estimates of seafloor topography. Nevertheless, it should be emphasized that such data products are indeed estimates, and that they may be inappropriate for certain types of geophysical modeling.

Global Earth topography models that combine oceanic bathymetry and landmass topography include ETOPO2 and SRTM30_PLUS. The later of these represents a combination of the SRTM 30-arcsec data and the Smith and Sandwell (1997) predicted bathymetry, with gaps filled by GTOPO30 data, among others. An image of this global topographic model is displayed in Figure 1. This data set represents

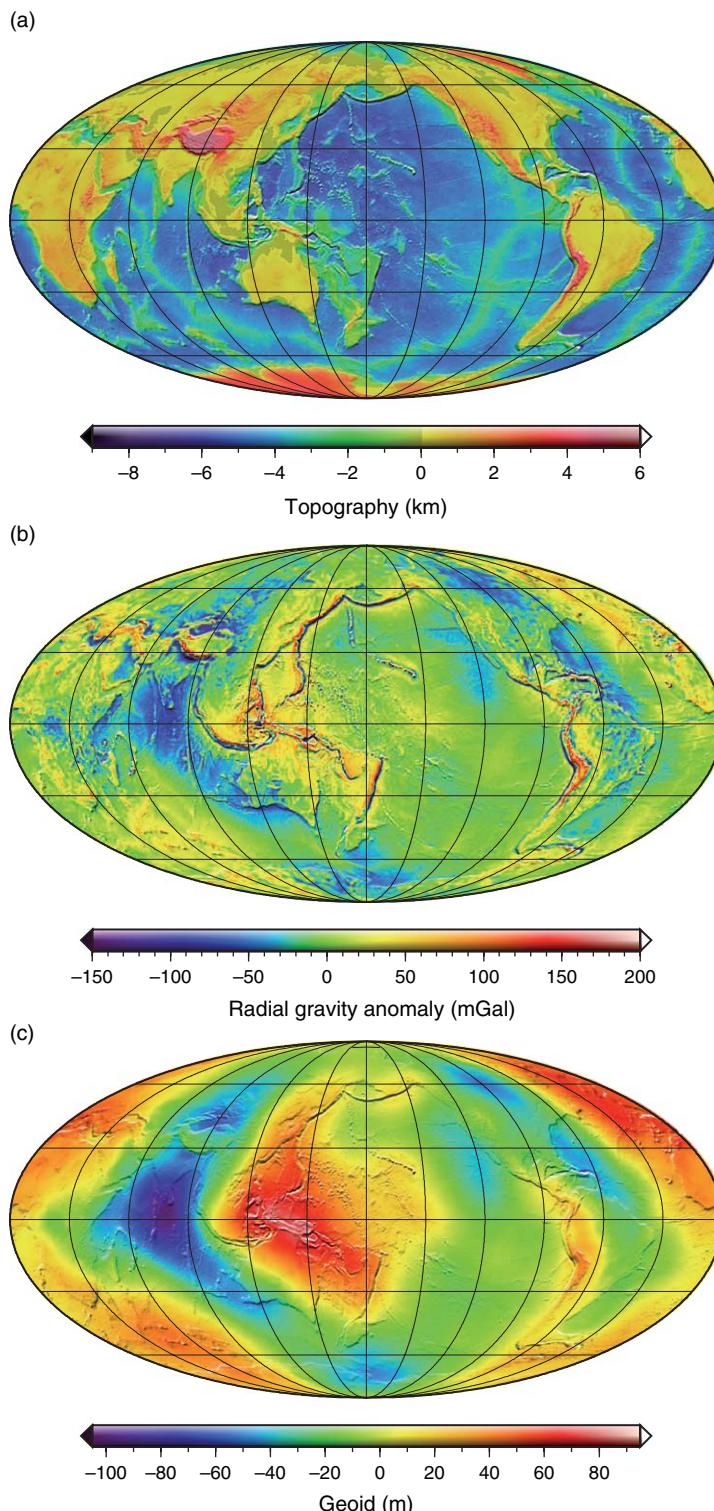


Figure 1 (a) Global topography and bathymetry of the Earth, referenced to mean sea level, of the model SRTM30_PLUS. (b) Radial free-air gravity from EIGEN-GL04C, obtained after setting the zonal degree-2 term equal to zero, evaluated at a radius of 6378.1 km. (c) First-order approximation to the geoid obtained from the same coefficients as the radial gravity field. All images are in a Mollweide projection with a central meridian of 180° W longitude and are overlaid by a gradient image derived from the topographic model.

elevations above the ‘WGS84 EGM96 geoid’, which is a good approximation of ‘mean sea level’ (see internet documentation in **Table 1** for precise definitions). If absolute radii of the Earth were desired (as is necessary for most methods that calculate gravity from topography), it would be necessary to add this geoid, which is referenced to the WGS84 ellipsoid, and the WGS84 ellipsoid itself. The WGS84 ellipsoid is a good representation of the Earth’s zonal shape, and possesses a 21 km rotationally induced difference in elevation between the equator and polar axis. A summary of fundamental planetary constants is given in **Table 2**.

10.05.3.1.2 Gravity

The gravitational field of the Earth has been mapped by several techniques, including analyses of satellite tracking data, terrestrial measurement campaigns, and satellite altimetry of the ocean surface (geoid slopes are proportional to the gravity field in the spectral domain (see, among others [Hwang and Parsons, 1996](#); [Sandwell and Smith, 1997](#))). The construction of global high-resolution models generally consists of combining the long-wavelength information from satellite tracking data with the short wavelength information in the terrestrial and oceanic altimeter surveys. The model EGM96 ([Lemoine et al., 1998](#)) has been the standard reference for much of the past

decade, but this has been superseded by data obtained from the recent missions CHAMP and GRACE.

In contrast to the EGM96 model that is based upon satellite tracking data from terrestrial stations, the ongoing missions CHAMP ([Reigber et al., 2004](#)) and GRACE ([Tapley et al., 2004](#)) are based upon continuous satellite to GPS (global positioning system) tracking data. While CHAMP is a single satellite, GRACE consists of two satellites on identical orbits of which the distance between the two is measured to high accuracy by a microwave communication link. Two of the most recent high resolution models of the Earth’s gravitational field derived from these data include GGM02C ([Tapley et al., 2005](#)) and EIGEN-GL04C (cf. [Reigber et al., 2005](#)). GGM02C is based upon GRACE tracking data combined with terrestrial and altimeter based surface measurements and is valid to degree 200 (this can be augmented to degree 360 by using the EGM96 spherical harmonic coefficients). In contrast, EIGEN-GL04C is valid to degree 360 ($\lambda \approx 110$ km) and additionally makes use of laser ranges to the LAGEOS satellites.

Images of the radial gravitational anomaly and geoid, as determined from eqns [21] and [22], respectively, are shown in **Figure 1** for the model EIGEN-GL04C. For both images, the zonal degree-2 term that is primarily a result of the rotational flattening was set to zero, and the fields were

Table 2 Gravitational and shape constants.

Parameter	Value	Source
G	$6.674\ 2(10)\ 10^{-11}\ m^3\ kg^{-1}\ s^{-2}$	Committee on Data for Science and Technology (Mohr and Taylor, 2005)
Earth		
GM	$398.600\ 441\ 5\ 10^{12}\ m^3\ s^{-2}$	EGM96; Lemoine et al. (1998)
Semi-major axis	6378.137 km	WGS84; National Imagery and Mapping Agency (2000)
Semi-minor axis	6356.752 3142 km	WGS84; National Imagery and Mapping Agency (2000)
Radius of sphere of equal volume	6371.000 79 km	WGS84; National Imagery and Mapping Agency (2000)
ω	$72.921\ 150\ 10^{-6}\ rad\ s^{-1}$	WGS84; National Imagery and Mapping Agency (2000)
Venus		
GM	$324.858\ 592\ 10^{12}\ m^3\ s^{-2}$	MGNP180U; Konopliv et al. (1999)
Mean planetary radius	6051.878 km	VenusTopo719
ω	$-299.24\ 10^{-9}\ rad\ s^{-1}$	Konopliv et al. (1999)
Mars		
GM	$42.828\ 374\ 568\ 10^{12}\ m^3\ s^{-2}$	JGM95I01; Yuan et al. (2001)
Mean planetary radius	3389.500 km	MarsTopo719
ω	$70.882\ 182\ 8\ 10^{-6}\ rad\ s^{-1}$	Yuan et al. (2001)
The Moon		
GM	$4.902\ 801\ 076\ 10^{12}\ m^3\ s^{-2}$	LP150Q; Konopliv et al. (2001)
Mean planetary radius	1737.064 km	MoonUSGS359
ω	$2.661\ 707\ 3\ 10^{-6}\ rad\ s^{-1}$	Yoder (1995)

evaluated at a reference radius of 6378.1 km. (The shape and potential of a flattened ellipsoid are well approximated by the degree-2 zonal harmonic.) The largest gravitational anomalies are seen to be correlated with topography (such as trenches and seamounts), and the geoid height is found to vary by about 200 m. Errors in the geoid and gravity field are estimated to be approximately 30 cm and 8 mGal, respectively.

10.05.3.1.3 Spectral analysis

A spherical harmonic model of the Earth's shape was constructed by adding the EGM96 and WGS84 geoid and ellipsoid, respectively, to the model SRTM30_PLUS, and expanding this gridded data set to degree 719 using the method of Driscoll and Healy (1994). For the analyses presented here, the radius of the ocean floor was increased by 1030/2670 times its depth in order to convert the mass of the overlying sea water into 'rock equivalent topography' (this model is here designated as SRTMP719 RET). Spectral and cross-spectral properties of this model and the EIGEN-GL04C potential are plotted in Figure 2. As demonstrated in the left panel, the power spectrum of the geoid is about 5 orders of magnitude less than that of the topography, which is a reflection of the low amplitudes of the geoid undulations present in Figure 1. The (calibrated) error spectrum of the potential model demonstrates that the coefficients are well determined at low degrees, with uncertainties gradually increasing to a near constant value close to degree 100. As a result of the \sim 400–500 km altitude of the GRACE and CHAMP satellites, the contribution to the gravitational field from the orbital tracking data is

necessarily limited to degrees less than about 100; the higher-degree terms are constrained almost entirely by the surface measurements.

The admittance and correlation spectra between the radial gravity and topography are plotted in the right panel of Figure 2. The correlation for many of the lowest degrees is seen to be small, and in some cases negative. Beyond degree 12 the correlation is relatively constant with a value of \sim 0.6–0.7, though it should be noted that this function slightly decreases with increasing degree beyond degree \sim 250. If the surface topography were completely uncompensated, which should be a good approximation beyond degree 200, then the admittance would have a near-constant value of $2\pi\rho G$, or 42 mGal km^{-1} times the surface density in units of g cm^{-3} . The observed function is everywhere lower than this theoretical value by a factor of two, and this is a consequence of the fact that the gravity and topography are not perfectly correlated on a global scale (see discussion in Section 10.05.6.2).

10.05.3.2 Venus

10.05.3.2.1 Topography

The planet Venus possesses a dense atmosphere and is perpetually enshrouded by opaque clouds of sulphuric acid. In order to obtain measurements of the surface, it is necessary to make use of electromagnetic frequencies, such as microwaves, where the atmosphere is transparent. Surface elevations of Venus have been measured from orbit using radar altimeters onboard the missions Pioneer Venus Orbiter, Venera 15 and 16, and Magellan. Of these, the Magellan spacecraft, which was in orbit between 1990 and

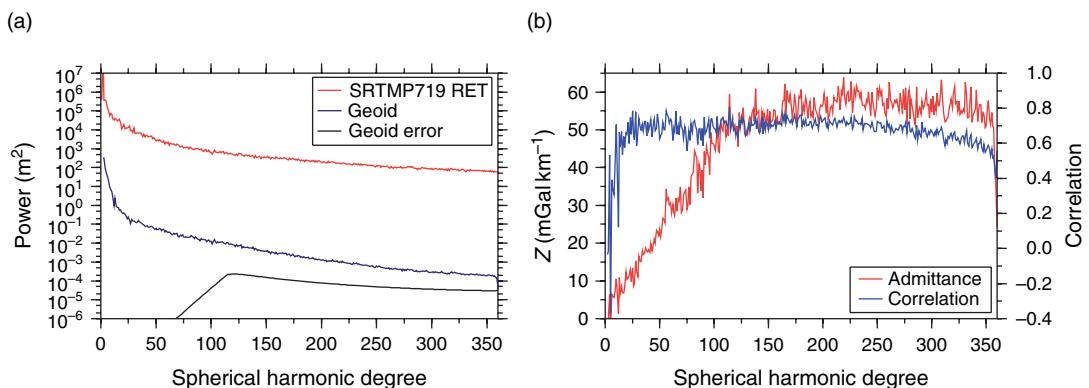


Figure 2 Power and cross-power spectra of the Earth's radial gravity and topography. (a) Power spectra of the topography (SRTMP719 RET), geoid (EIGEN-GL04C), and calibrated geoid error. (b) Admittance and correlation spectra of the radial gravity and topography.

1994, collected the highest resolution measurements on a near-global scale (for a detailed description, see Ford and Pettengill, 1992).

As a result of the elliptical orbit of the Magellan spacecraft, the spatial resolution of the elevation measurements varied between 8×10 km at periape to 19×30 km at the North Pole (Rappaport *et al.*, 1999). Over 4 million range measurements were ultimately collected, and these were used to construct a 5×5 km gridded elevation model. With the exception of a few relatively minor data gaps covering about 2.6% of the planet's surface area, coverage of the planet is fairly uniform. (If data from previous missions are used to fill the gaps, then just less than 1% of the planet is not covered at this resolution.) Though the spacecraft-surface range measurements are estimated to have an accuracy of less than 10 m (errors can be larger over rough and steeply sloping terrain), uncertainties in the spacecraft orbit at the time of initial processing were sometimes much greater, especially during superior conjunction. The most recent gravity model of Konopliv *et al.* (1999) has considerably improved the spacecraft navigational errors, and these improved orbit predictions have been used by Rappaport *et al.* (1999) in a complete reprocessing of the altimetry data set (archived as GTDR3.2). Horizontal uncertainties in the footprint locations are insignificant in comparison to the footprint size, and the RMS radial uncertainty is estimated to be less than 20 m.

An image of the Venusian topography (derived from the spherical harmonic model VenusTopo719; see below) is shown in **Figure 3**, where it is referenced to the geoid. While the hypsometry of Venus is unimodal (e.g., Ford and Pettengill, 1992; Rosenblatt *et al.*, 1994), in contrast to that of the Earth which is bimodal, Venus can be broadly characterized by its low-lying plains, ‘continental’ plateaus, and volcanic swells. The most prominent highlands include Aphrodite Terra, which lies along the equator, and Ishtar Terra, which is located at high northern latitudes. Ishtar and Aphrodite Terra differ in that the former is flanked by high elevation mountain ranges. Isolated domical volcanic provinces that possess prominent rift valleys include Atla (0° N, 200° E) and Beta (25° N, 280° E) Regiones. The highest topographic excursion corresponds to Maxwell Montes, located in Ishtar Terra, which reaches more than 10 km above the surrounding plains. (Maps with feature names for Venus, Mars, and the Moon can be found at the appropriate web address in **Table 1**.)

10.05.3.2.2 Gravity

Models of the gravitational field of Venus have been constructed through the analyses of tracking data from the Pioneer Venus Orbiter and Magellan spacecraft (for a review, see Sjogren *et al.*, 1997). The orbit of the Pioneer Venus spacecraft was highly eccentric, and possessed periape altitudes as low as 150 km near the equator. The Magellan spacecraft was initially on an eccentric orbit as well, but through the technique of aerobraking during the gravity mapping phase of the mission, the orbit was transformed to a more circular one, with periape and apoapse altitudes varying between 155–220 and 350–600 km, respectively.

The most recent model of the Venusian gravitational field is the degree 180 JPL (Jet Propulsion Laboratory) model MGNP180U of Konopliv *et al.* (1999). Because of computational constraints at the time, this model was constructed in three phases. In the first step, a model to degree 120 was generated using the full unconstrained covariance matrix and a spatial *a priori* constraint that depended on the strength of the gravitational accelerations (such models are labeled SAAP for Surface Acceleration *A Priori*). The second step used this model as the nominal solution, and then solved for the coefficients from degree 116 to 155 using the same spatial constraint. For the third step, the coefficients were determined from degree 154 to 180, but instead of using a spatial constraint, the spherical harmonic coefficients were biased towards a global power law (i.e., a ‘Kaula rule’). Future models could be improved by performing the inversion in a single step. As a result of the spatial constraint that was employed in the first two steps, the spatial resolution of the model varies dramatically with position on the surface. While spectral resolutions approaching degree 180 may be realized close to the equator, other regions possess resolutions as low as degree 40 (see figure 3 of Konopliv *et al.*, 1999).

Images of the MGNP180U radial gravitational field and geoid are presented in **Figure 3**, evaluated at a radius of 6051 km, where the spectral coefficients have been truncated beyond degree 65. As a result of the slow retrograde rotation of Venus, there is no appreciable rotational flattening of the planet, and the J_2 potential coefficient is thus here included in contrast to that of the Earth and Mars. These plots show that most gravitational and geoid anomalies are highly correlated with the surface topography. The largest radial gravitational anomalies are associated with the volcanoes Maat and Ossa Mons in Atla

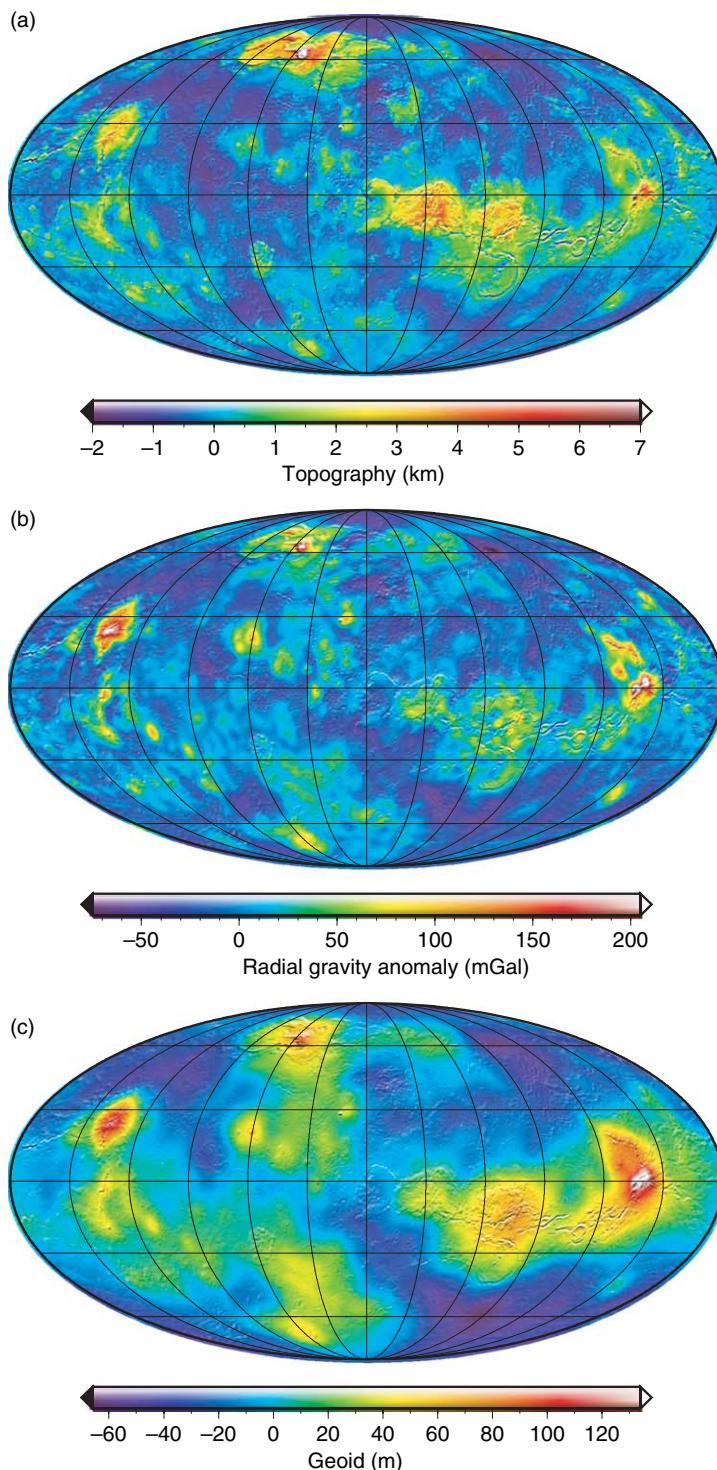


Figure 3 (a) Global topography of Venus, derived from the spherical harmonic model VenusTopo719, referenced to the geoid. (b) Radial free-air gravity, evaluated at a radius of 6051 km, obtained after truncating the spherical harmonic coefficients of MGNP180U beyond degree 65. (c) First-order approximation to the geoid obtained from the same coefficients as the radial gravity field. All images are in a Mollweide projection with a central meridian of 60° E longitude and are overlaid by a gradient image derived from the topographic model.

Regio, with values reaching about 270 mGal. The high elevations of Maxwell Montes, Beta Regio, and numerous smaller volcanic provinces, are also seen to possess significant anomalies. Uncertainties in the radial component of the gravitational field are typically 10 mGal at the surface, but can be as high as 50 mGal in places.

Like the Earth, the geoid undulations of Venus possess a dynamic range of only \sim 200 m. The largest geoid anomalies correspond to the volcanic swells of Atla and Beta Regiones, and the continental regions of Aphrodite and Ishtar Terra. It is also seen that the plains with the lowest elevations possess negative geoid anomalies. Uncertainties in the geoid are typically 1 m, but can reach values as high as 4 m.

10.05.3.2.3 Spectral analysis

A degree 719 spherical harmonic model (VenusTopo719) was constructed based on the Magellan GTDR3.2 sinusoidally projected data product. Missing nodes were filled by data obtained by the Pioneer Venus and Venera 15/16 missions, the remaining gaps were filled by interpolation using the GMT (Wessel and Smith, 1991) command ‘surface’ with a tension parameter of 0.35, and the spherical harmonic expansion was performed using the sampling theorem of Driscoll and Healy (1994). While the resulting power spectrum was found to be insensitive to changes in the tension parameter, the mean planetary radius varied by about 1 m among the various models. A comparison between this spherical harmonic model and the one of Rappaport *et al.* (1999) shows that the latter suffers from an increasing loss of fidelity with increasing degree (the degree correlation between the two models is \sim 0.93 at

degree 360). While the cause for this is uncertain, possible explanations might be the use of inaccurate Legendre functions or the presence of short-wavelength aliases at high degrees.

Power spectra of the Venusian topography (VenusTopo719) and geoid (MGNP180U) are shown in the left panel of Figure 4. These are similar to those of the Earth, with the exception that the amplitudes of the degree-1 and -2 topographic terms for Venus are relatively smaller. On a log–log plot (not shown), a change in slope of the topographic power spectrum is seen to occur near degree 100 (Rappaport *et al.*, 1999). While this feature might be real, it is also possible that it is related to interpolating over data gaps before performing the spherical harmonic expansion. The error spectrum of the geoid is seen to be greater than the geoid itself for degrees greater than 65. While the global values of the potential coefficients are generally unreliable beyond this degree, it should be noted that the spatial resolution of the gravitational field is a strong function of position on the surface. Discontinuities in the error spectrum are artifacts of solving for the potential coefficients in three separate steps.

The spectral admittance and correlation functions for the radial component of the gravitational field and topography, plotted in the right panel of Figure 4, are seen to differ significantly from those for the Earth. The admittance is found to possess values between \sim 30 and 50 mGal km $^{-1}$ for degrees up to 100, whereas for the Earth, the admittance linearly increases from \sim 0 to 30 mGal km $^{-1}$ at degree 100. The correlation between the gravitational field and topography is also significantly higher for degrees less than 40 than it is for the

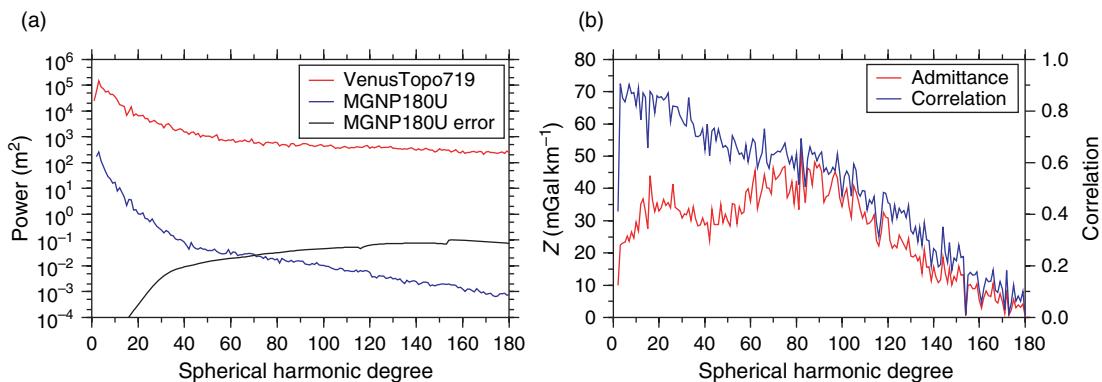


Figure 4 Power and cross-power spectra of the gravity and topography of Venus. (a) Power spectra of the topography (VenusTopo719), geoid (MGNP180U), and geoid error. (b) Admittance and correlation spectra of the radial gravity and topography.

Earth. Nevertheless, beyond degree 60, the spectral correlation and admittance are seen to linearly decrease with increasing degree, which is a result of the poor determination of the global potential coefficients. It is of note that both the admittance and correlation for degree 2 are significantly smaller than the neighboring values. As these are unaffected by the slow retrograde rotation of Venus, these low values may demand an origin that is distinct from the higher degrees.

Because the Pioneer Venus Orbiter and Magellan spacecraft were in near polar orbits, the gravity field is better determined for the near-sectoral terms. Sectoral terms correspond to when $|m|=l$, for which the corresponding spherical harmonic functions do not possess any latitudinal zero crossings. By considering only those coefficients where $l-|m|<20$, Konopliv *et al.* (1999) have shown that both the admittance and correlation between gravity and topography are considerably greater than when considering all coefficients combined. In particular, the correlation function remains close to 0.7 for degrees up to 140, at which point it decreases substantially. Thus, while high-degree localized spectral analyses may be justified on Venus, the fidelity of the spectral estimates will be a strong function of both position and the spherical-harmonic degree and order.

10.05.3.3 Mars

10.05.3.3.1 Topography

Prior to the 1990s, the best Martian topographic models were constructed by a combination of Earth-based radar data, spacecraft radio occultations, and stereo and photoclinometric observations, all of which suffered from either large uncertainties or a limited spatial extent (for a review, see Esposito *et al.*, 1992). The laser altimeter onboard the MGS spacecraft (MOLA; Mars Orbiting Laser Altimeter) has since collected an impressive data set that has revolutionized studies of the Martian surface (see Smith *et al.*, 1999, 2001b; Zuber *et al.*, 2000a).

After being inserted into orbit in 1997, MOLA made more than 640 million ranges to the surface over the period of 4 years. The spot size of the laser at the surface was ~ 168 m, and these were spaced every 300 m in the along-track direction of the spacecraft orbit. The intrinsic range resolution of the instrument was 37.5 cm, but range precision decreases with increasing surface slope, and could be as poor as 10 m for slopes near 30° . While the along-track

orbit errors are less than the size of the laser footprint, radial orbit errors could sometimes be as high as 10 m. Nevertheless, these orbit-induced errors could be minimized by the use of altimetric crossovers. Crossovers occur whenever two altimeter ground tracks intersect, and the difference in the two elevation measurements is largely a reflection of errors in the orbit determination. By parametrizing these uncertainties by a slowly varying function, the crossover residuals can be minimized (Neumann *et al.*, 2001). Such a procedure was capable of reducing the rms crossover residuals from about 8.3 to 1.8 m. Using these methods, it has been possible to measure temporal variations in CO₂ snow depth that can reach 2 m in the polar regions (Smith *et al.*, 2001a).

The topography of Mars (as determined from the 719-degree spherical harmonic model MarsTopo719; see below) referenced to the geoid (calculated to second-order accuracy) is displayed in Figure 5. Two of the most remarkable features are the dichotomy in elevation between the Northern and Southern Hemispheres, and the regionally high elevations of the Tharsis volcanic province that is centered near the equator at 100° W. These two features give rise to a 3.3 km offset of the center of figure from the center of mass that is directed toward 64° S and 99° W, of which the longitudinal offset is directed towards the Tharsis province. In addition to these long wavelength features, there is also an ~ 20 km difference between the polar and equatorial radii that is principally a result of the planet's rotational flattening.

Other major topographic features include the giant impact basins Hellas (40° S, 65° W), Argyre (50° S, 40° W) and Isidis (15° N, 85° E), the Elysium volcanic province (25° N, 145° E), the rift valley Valles Marineris, and the prominent volcanoes that are superposed on the Tharsis province. The highest elevation corresponds to the volcano Olympus Mons, which rises almost 22 km above the MOLA reference geoid.

10.05.3.3.2 Gravity

The gravitational field of Mars has been successively improved by tracking data obtained from the Mariner 9, Viking 1 and 2, MGS, and Mars Odyssey missions. A major improvement in the gravity models came with the acquisition of data from the MGS mission (see Yuan *et al.*, 2001; Lemoine *et al.*, 2001; Konopliv *et al.*, 2006). This spacecraft was in a near-polar orbit, and during the early portion of the mission when the orbit was highly elliptical, tracking

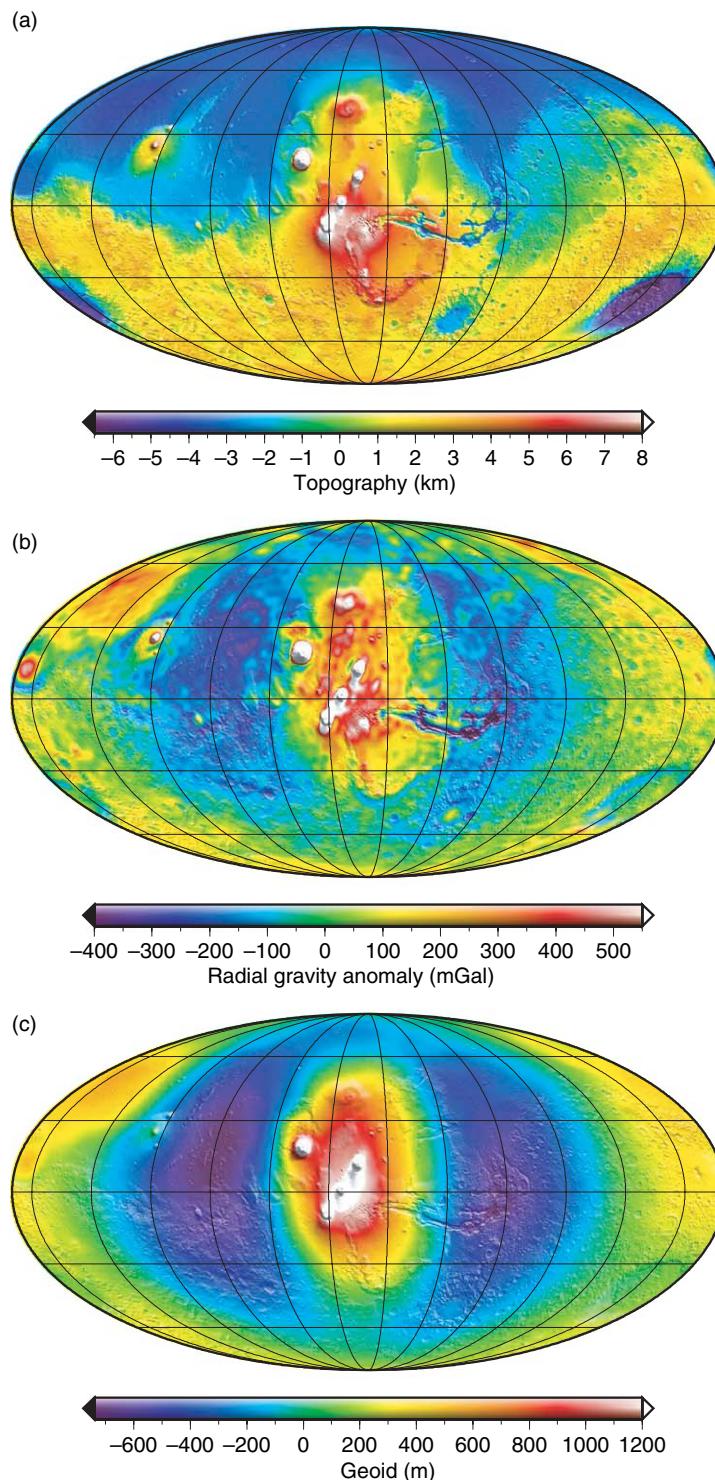


Figure 5 (a) Global topography of Mars derived from the 719-degree spherical harmonic model MarsTopo719, referenced to the geoid calculated to second order accuracy. (b) Radial free-air gravity, evaluated at a radius of 3396 km, obtained after truncating the spherical harmonic coefficients of JGM95J01 beyond degree 75 and setting the J_2 term to zero. (c) First-order approximation to the geoid obtained from the same coefficients as the radial component of the gravitational field. All images are in a Mollweide projection with a central meridian of 100° W longitude and are overlaid by a gradient image derived from the topography model.

data from altitudes as low as 170 km were acquired at latitudes between 60° and 85° N. Through the technique of aerobraking, the spacecraft was put into a near-circular mapping orbit with periapse and apoapse altitudes of 370 and 435 km, respectively.

The most recent and highest resolution gravity model of Mars is the JPL degree 95 model JGM95J01 (Konopliv *et al.*, 2006). This model employs MGS and Mars Odyssey tracking data in combination with surface tracking data from the Pathfinder and Viking 1 landers, the latter of which were used to improve knowledge of the orientation of Mars. Some of the Mars Odyssey tracking data are from altitudes as low as \sim 200 km (primarily over the Southern Hemisphere), and aerobraking eventually put the spacecraft into a near-circular 390 × 455 km mapping orbit. In the absence of *a priori* constraints, inversions for the global spherical-harmonic coefficients give rise to an unrealistic power spectrum beyond degree 60. In order to obtain a higher resolution model with realistic power, the JGM95J01 model was biased towards an *a priori* power law for degrees greater than 59 by use of a Kaula rule.

An image of the radial component of the JGM95J01 gravity model is plotted in Figure 5, where the spherical-harmonic coefficients have been truncated beyond degree 75 and the \mathcal{J}_2 term has been set to zero. (A small portion of the \mathcal{J}_2 potential is likely related to the nonhydrostatic mass anomaly associated with the Tharsis bulge (e.g., Zuber and Smith, 1997).) Clearly visible are the large positive anomalies associated with the volcanoes in the Tharsis plateau, and a negative gravity moat that surrounds this plateau (Phillips *et al.*, 2001).

Large positive anomalies are also evident for some of the largest impact basins, such as Isidis, Argyre, and the buried Utopia basin (45° N, 110° E) that lies in the northern plains. A negative gravitational anomaly is clearly associated with the Valles Marineris rift valley, and negative anomalies surrounding some mountains and volcanoes seem to indicate a flexural origin. Formal uncertainties in the radial gravity are at most 10 mGal (A. Konopliv, personal communication).

The Martian geoid, as obtained from the model JGM95J01, is shown in Figure 5. The geoid undulations of Mars (after removal of the \mathcal{J}_2 term) are seen to possess the largest amplitudes among the terrestrial planets, with a dynamic range of over 2.5 km. The signal is clearly symmetric about the Tharsis province, where a central geoid high is surrounded by an annular low. Other geoid highs are associated with the impact basins Isidis and Utopia, as well as the Elysium volcanic rise. Uncertainties in the geoid are generally no more than 2 m (A. Konopliv, personal communication).

10.05.3.3.3 Spectral analysis

The power spectra of the Martian topography (MarsTopo719; calculated from the gridded data sets available at the PDS website) and geoid (JGM95J01) are plotted in Figure 6. In comparison to the topographic power spectrum, the Martian geoid is seen to have greater amplitudes by about two orders of magnitude than both the Earth and Venus. Furthermore, the first 5 degrees of the Martian geoid are considerably greater than would be expected based upon an extrapolation of the higher degree terms. This low-degree signal is likely a consequence

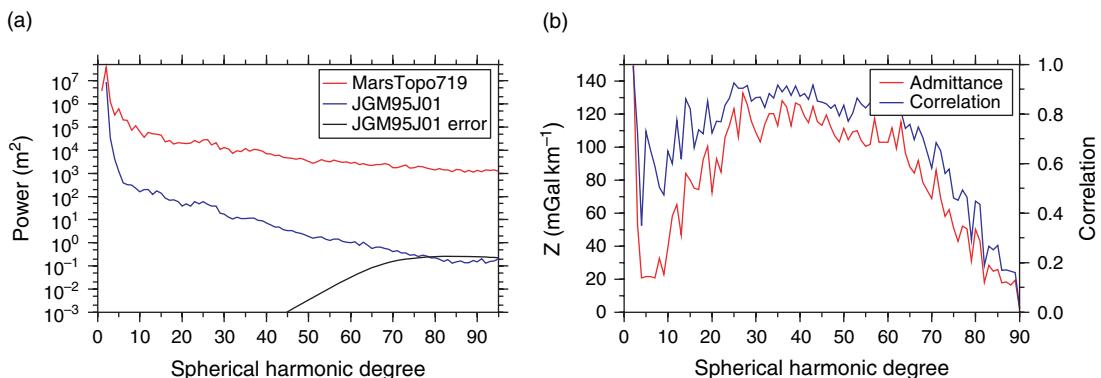


Figure 6 Power and cross-power spectra of the gravity and topography of Mars. (a) Power spectra of the topography (MarsTopo719), geoid (JGM95J01), and geoid error. (b) Admittance and correlation spectra of the radial gravity and topography.

of the lithospheric load and deflection associated with the Tharsis province (see Zuber and Smith, 1997; Phillips *et al.*, 2001). The error spectrum of the geoid is seen to be larger than the signal for degrees greater than ~ 75 .

The admittance and correlation of the radial gravitational field and topography are shown in the right panel of **Figure 6**. The admittance function gradually increases with increasing degree, attaining a relatively constant value beyond degree 30. Beyond degree 65 both the admittance and correlation decrease as a result of the poor resolution of the gravity field. While the shape of the admittance function is somewhat similar to that of the Earth, it is important to note that the amplitudes are considerably larger at high degrees (~ 100 in comparison to $\sim 35 \text{ mGal km}^{-1}$). Indeed, these large values are comparable to what would be expected for uncompensated topography. One apparent anomaly with the admittance spectrum is the relatively high value of 53 mGal km^{-1} for the degree three term.

With few exceptions (such as degrees 4 and 9) the correlation between the gravity and topography is also seen to be very high, with values between about 0.6 and 0.8. Similar to the Venusian gravity model, the near-sectoral terms of the Martian gravity solutions are relatively better determined because of the near-polar orbit of the MGS spacecraft. When only these near-sectoral terms are used, the correlation between the gravity and topography is considerably larger at high degrees in comparison to the case shown in **Figure 6** (see Yuan *et al.*, 2001; Konopliv *et al.*, 2006).

10.05.3.4 The Moon

10.05.3.4.1 Topography

The topography of the Moon has been measured by several means, including satellite altimetry, stereo-photogrammetry, and radar interferometry (see Wieczorek *et al.*, 2006, for a more detailed review). However, because of the Moon's synchronous rotation, most early studies were restricted either to the nearside hemisphere, or along the equatorial ground tracks of the Apollo command and service modules. While pre-Apollo stereo-photogrammetric studies were successful in obtaining regional topographic models with good relative precision, the long wavelength and absolute accuracies of these models were much poorer.

The Clementine mission, launched in 1994, was the first to measure absolute elevations of the lunar

surface on a near global scale (see Zuber *et al.*, 1994; Smith *et al.*, 1997). North-south topographic profiles were obtained between 79°S and 22°N during the first month of this mission, and then between 20°S and 81°N during the second. The absolute accuracies of the obtained surface elevations are about 100 m, the cross-track orbit spacing was about 60 km at the equator, and the minimum along-track shot spacing was about 20 km (1° at the equator corresponds to 30 km). As a result of the nonoptimal design of the lidar, however, the electronics often detected many returns, and these needed to be filtered to determine which, if any, were from the lunar surface. The returns from many shots were ultimately discarded, especially over the rougher highlands, leading to the acceptance of a total of 72 548 range measurements. Comparisons with a radar interferometry-derived topographic model of the crater Tycho (Margot *et al.*, 1999a) suggests that a few percent of the accepted Clementine range measurements are erroneous.

As a result of the polar orbit of the Clementine spacecraft, many overlapping images exist in the polar regions under varying viewing conditions. These have been used to construct regional digital elevation models poleward of 60° having a 1-km spatial resolution (Cook *et al.*, 2000; US Geological Survey, 2002). While the relative elevations obtained from these studies were tied to the Clementine altimeter data near the outer edges of these models, absolute accuracy is expected to degrade towards the poles. In particular, comparisons with independent regional models of the polar regions based on radar interferometry data (Margot *et al.*, 1999b) show differences of a kilometer or more.

The US Geological Survey (2002) topographic model represents a combination of interpolated Clementine altimetry and elevation models of the polar regions based on stereo photogrammetry. This model is presented in **Figure 7** where it is referenced to the geoid, which includes the static gravitational model LP150Q and the rotational and tidal contributions of eqn [18]. The most dramatic feature of the Moon's topography is seen to be the giant South Pole-Aitken impact basin on the southern farside hemisphere. This impact basin possesses a total relief of over 10 km, and with a diameter of over 2000 km, it is the largest recognized impact structure in the solar system. Other impact basins and craters of various sizes are seen to have shaped the relief of the lunar surface, and the extensive mare basaltic lava flows on the nearside hemisphere, which are relatively younger,

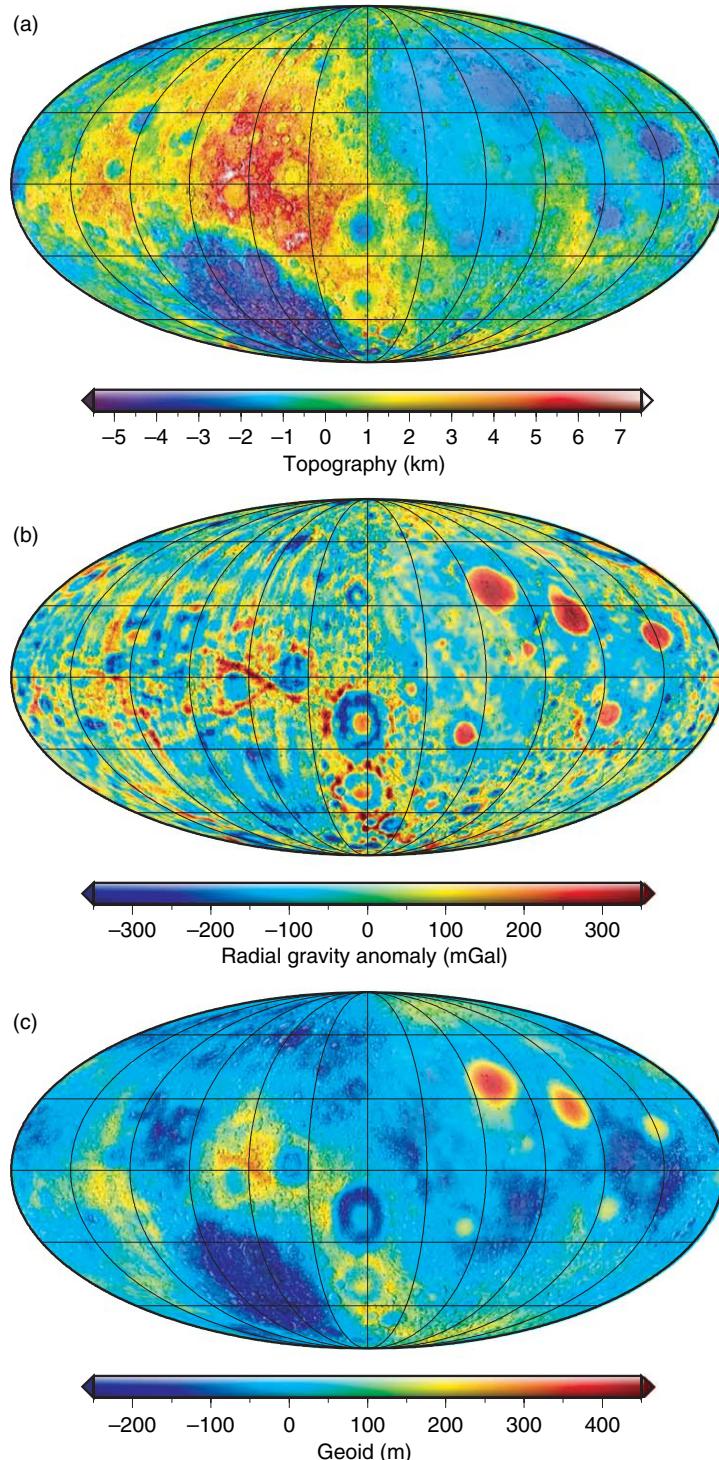


Figure 7 (a) Global topography of the Moon from the spherical harmonic model MoonUSGS359 referenced to the geoid which includes both the LP150Q gravitational model and rotational and tidal effects. (b) Radial free-air gravity obtained from the model LP150Q evaluated at a radius of 1738 km after truncating the coefficients beyond degree 130 and setting the C_{20} and C_{22} terms equal to zero. (c) First-order approximation to the geoid obtained from the same coefficients as the radial gravity field. All images are in a Mollweide projection with a central meridian of 90° W longitude and are overlaid by the shaded relief map of Rosiek and Aeschliman (2001). The near- and farside hemispheres are on the right and left halves of these images, respectively.

are seen to be comparatively smooth. Also of note is that the Moon possesses a 1.9 km displacement of its center of figure from its center of mass in the direction of 8° N and 157° W (e.g., [Smith et al., 1997](#)).

10.05.3.4.2 Gravity

The gravitational field of the Moon has been determined by analyses of radio tracking data of orbiting spacecraft, which include data from the Lunar Orbiter, Apollo, Clementine, and Lunar Prospector missions. While all data contribute to the lunar gravity models, by far the highest resolution constraints are obtained from the extended Lunar Prospector mission when the spacecraft altitude was lowered to approximately 30 km above the surface (for a detailed discussion, see [Konopliv et al., 2001](#)). However, despite this low-altitude tracking data, because of the Moon's synchronous rotation, global models of the gravitational field are severely hindered by the lack of tracking data over the lunar farside.

While spacecraft have been tracked approximately 20° over the lunar limb, there is a sizable portion of the lunar surface that lacks direct tracking constraints. Regardless, as the long term orbits of lunar satellites are influenced by gravitational anomalies that are present there, some information can be extracted over these regions when inverting the tracking data. When no *a priori* constraints are used in constructing the gravity model, the field is found to be completely unreliable in an approximately 60° radius 'shadow zone' centered on the antipode of the sub-Earth point. Globally, such unconstrained models are only reliable to spherical-harmonic degree 15. In order to obtain solutions with reasonable characteristics, it is necessary to bias the spherical-harmonic coefficients towards an *a priori* power spectrum (i.e., by use of a 'Kaula rule'). Using such methods, the most recent JPL gravity model LP150Q ([Konopliv et al., 2001](#)) has been determined to degree 150. Attempts to obtain regional models on the nearside with a higher resolution can be found in [Goossens et al. \(2005a\)](#).

The LP150Q radial gravity field of the Moon (truncated at degree 130) is plotted in [Figure 7](#). As the equilibrium shape of a synchronously locked satellite contains both C_{20} and C_{22} terms, both of these have been removed. (While these two terms are small at the current Earth-Moon separation, it has been suggested that the Moon may possess a fossil shape that was frozen into the lithosphere early in its orbital evolution (e.g., [Jeffreys, 1976](#); [Lambeck and Pullan, 1980](#).) The major features of this map

include the large positive gravitational anomalies associated with the nearside impact basins, colloquially referred to as 'mascons' (e.g., [Muller and Sjogren, 1968](#)), negative gravity moats that surround some of these basins, and the more noisy and less constrained farside field. Despite the lack of direct farside tracking data, it is remarkable that the inferred gravitational anomalies there correlate with large impact basins. Nevertheless, care should be used when interpreting these anomalies as their amplitudes could be muted, and/or their positions laterally offset. The uncertainties in the radial component of the gravitational field are estimated to be approximately 30 mGal on the nearside and can reach up to 200 mGal on the farside.

A plot of the lunar geoid is also shown in [Figure 7](#) after removing both the C_{20} and C_{22} terms. In contrast to the Earth, which possesses maximum geoid undulations of ±100 m, the dynamic range of the lunar geoid is more than 700 m. When considering phenomena such as basalt flow directions, it is thus necessary to use elevations that are referenced to the full geoid. Uncertainties in the geoid are estimated to be approximately 4 m on the nearside and 60 m on the farside.

Finally, it is noted that the orientation of the Moon is completely described by its three principle moments of inertia, which in turn completely determine the degree-2 gravity coefficients (e.g., [Lambeck, 1988](#)). For a synchronously rotating satellite, the minimum energy configuration is achieved when the maximum moment of inertia lies along the rotation axis, and when the minimum moment coincides with the Earth-Moon direction. A 180° rotation of the Moon about its rotation axis would be equally stable as its current configuration.

10.05.3.4.3 Spectral analysis

A spherical harmonic model of the USGS topography was developed to degree 359 and is here designated as MoonUSGS359. The power spectrum of this model, the prior model GLTM2C ([Smith et al., 1997](#)) that is based solely on altimetry data, and the LP150Q geoid and error are plotted in the left panel of [Figure 8](#). Concerning the two topographic models, it is seen that the power spectra of these diverge near degree 25. Careful inspection of how these models were constructed indicates that this is a result of using different methods to interpolate the sparse Clementine data points. In comparison to the Earth, the power spectrum of the lunar geoid is seen to be at least an order of magnitude more important when

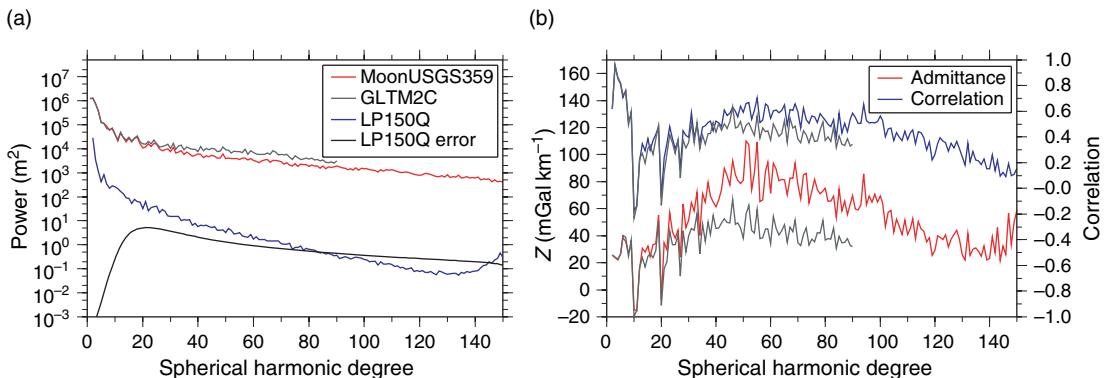


Figure 8 Power and cross-power spectra of the Moon’s gravity and topography. (a) Power spectra of the topographic models MoonUSGS359 (color) and GLTM2C (gray), and the LP150Q geoid and error. (b) Admittance and correlation spectra of the radial gravity and topography using MoonUSGS359 (color) and GLTM2C (gray).

compared to the topography spectrum. The upturn in the geoid spectrum beyond degree 130 is a result of short-wavelength aliasing in the gravity solution, and the error spectrum of the geoid is seen to be greater than that of the geoid itself beyond degree 80. While the global harmonic coefficients should be considered unreliable beyond this degree, it should be recognized that the uncertainty in the gravity field is a strong function of position.

The correlation and admittance spectra for the gravity and topography models are plotted in the right pane of Figure 8. The curves in color represent those obtained from the model MoonUSGS359, whereas the gray curves are for the model GLTM2C. Both correlation curves show that some of the lowest degrees of the radial gravity and topography are anticorrelated. These same degrees possess negative admittance values, and this is simply a result of the presence of ‘mascon’ impact basins, which are positive gravitational anomalies possessing low elevations. Beyond degree 25 it is seen that both the admittance and correlation spectrum obtained from the USGS model are somewhat greater than those from the GLTM2C model. While this is partially the result of the inclusions of high resolution polar topography in the USGS model, it is also likely that this is a consequence of the different interpolation schemes used in generating these models. As a decrease in correlation would be expected for a lower fidelity model, the USGS model should probably be preferred over GLTM2C. The decrease in the USGS admittance spectrum beyond degree 50 is likely to represent a loss of fidelity in both the gravitational and topographic models.

10.05.4 Methods for Calculating Gravity from Topography

Geophysical investigations that employ gravity and topography data often attempt to fit the observations with those predicted from a model that contain several parameters. For example, lithospheric flexure calculations depend upon several unknowns, including the effective elastic thickness of the lithosphere, crustal thickness, and the density of the crust and mantle. By comparing predicted gravitational anomalies induced by the deflection of density interfaces with the observed values, the parameters of such a model can be constrained. While several methods exist for calculating the gravitational field related to relief along a density interface, as is described below, this is oftentimes most easily performed in the spherical harmonic domain.

The calculation of the gravitational potential can be considerably simplified by use of the two identities:

$$\frac{1}{|\mathbf{r}-\mathbf{r}'|} = \frac{1}{r} \sum_{l=0}^{\infty} \left(\frac{r'}{r} \right)^l P_l(\cos \gamma) \quad \text{for } r \geq r' \quad [27]$$

$$P_l(\cos \gamma) = \frac{1}{2l+1} \sum_{m=-l}^l Y_{lm}(\theta, \phi) Y_{lm}(\theta', \phi') \quad [28]$$

where P_l is an unnormalized Legendre Polynomial, and γ is the angle subtended between \mathbf{r} and \mathbf{r}' . (Equation [28] is commonly referred to as the Legendre addition theorem.) By inserting these equations into the expression for the gravitational potential (eqn [14]), it is simple to show that the

spherical harmonic coefficients of eqn [16] (the ‘Stokes coefficients’) are equal to

$$C_{lm} = \frac{1}{MR_0^l(2l+1)} \int_V \rho(r') r'^l Y_{lm}(\theta', \phi') dV' \quad [29]$$

It is important to note that this formulation of the gravitational potential can only be used when the observation point is greater than the maximum radius of the body.

Next, consider the case where there is relief $b(\theta, \phi)$ referenced to a spherical interface of radius D , and where the density ρ between b and D depends only on latitude and longitude (when b is negative, ρ is considered negative as well). For this situation, it is possible to obtain exact expressions for the corresponding potential coefficients that are similar to those developed by Parker (1972) in the Cartesian domain. Integrating eqn [29] over r' , and expanding powers of the relief in a Taylor series, the potential coefficients of eqn [16], referenced to a radius D , can be shown to be (see Wieczorek and Phillips, 1998)

$$C_{lm} = \frac{4\pi D^3}{M(2l+1)} \sum_{n=1}^{l+3} \frac{(\rho b^n)_{lm}}{D^n n!} \frac{\prod_{j=1}^n (l+4-j)}{(l+3)} \quad [30]$$

The spherical harmonic coefficients of the density multiplied by the relief to the n th power have the explicit expression (cf. eqn [8])

$$(\rho b^n)_{lm} = \frac{1}{4\pi} \int_{\Omega} [\rho(\theta, \phi) b^n(\theta, \phi)] Y_{lm}(\theta, \phi) d\Omega \quad [31]$$

and when the density is constant, eqn [30] reduces to eqn [9] of Wieczorek and Phillips (1998). As a result of the inequality in the identity of eqn [27], this expression for the potential is only valid when the radius r is greater than the maximum elevation $D + b$. Extensions, special cases, and alternative forms of this equation have been derived independently several times in the literature (e.g., Chao and Rubincam, 1989; Martinec *et al.*, 1989; Rapp, 1989; Balmino, 1994; Chambat and Valette, 2005).

For the common case where the density ρ is constant, the potential coefficients can be obtained simply by calculating the spherical harmonic coefficients of the relief to the n th power. While the sum of eqn [30] is finite, and hence exact, the number of terms grows linearly with spherical harmonic degree. Nevertheless, as each succeeding term is smaller than the previous, in practice, this sum can be truncated beyond a maximum value n_{\max} for which the

truncated terms are smaller than the resolution of the gravity model.

For certain applications it is sometimes sufficient to use the first order term of eqn [30]:

$$C_{lm} = \frac{4\pi D^2 (\rho b)_{lm}}{M(2l+1)} \quad [32]$$

This expression is commonly referred to as the ‘mass-sheet’ approximation, as the calculated gravitational anomaly would be exact if it arose from a spherical interface with a surface density of ρb . (The higher-order terms are referred to as the ‘finite amplitude’ or ‘terrain’ correction.) Using this expression, the radial gravity (see eqn [22]) is seen to asymptotically approach with increasing l the Bouguer slab approximation of $2\pi\rho Gb$.

The effect of truncating the sum of eqn [30] beyond n_{\max} is illustrated in Figure 9 for the specific case of determining the Bouguer correction of the Earth, Venus, Mars, and the Moon. The term Bouguer correction here refers to the contribution of the gravitational potential that results from the mass between the mean planetary radius and the surface. The true value of the Bouguer correction was approximated using $n_{\max} = 10$, and the maximum difference in the space domain that results from truncating at lower values of n was calculated on a spherical surface corresponding to the maximum radius of the planet. As is seen, in order to obtain accuracies of a few mGal, at least the first three terms

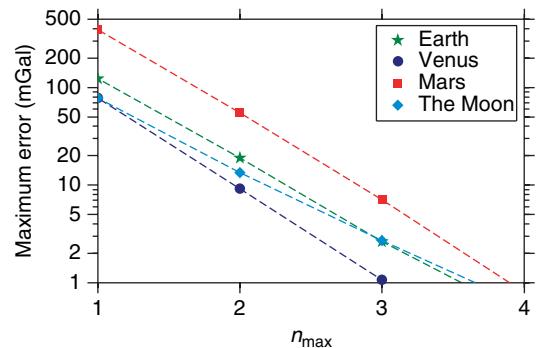


Figure 9 Maximum error associated with the Bouguer correction as a function of the order n_{\max} used in eqn [30]. The Bouguer correction is here defined as the radial gravitational attraction resulting from surface relief with respect to the mean planetary radius. The true value was approximated by $n_{\max} = 10$, the crustal density was assumed to be 2900 kg m^{-3} , and the radial gravitational anomaly was evaluated at the maximum elevation of the planet.

of eqn [30] are required. Utilizing only the first-order term could incur errors of a few hundred mGal for regions with high elevations.

Finally, it is noted that alternative means exist for calculating the theoretical gravitational field of a body, and that these may be preferable to the above approach for certain applications. One method developed by [Belleguic et al. \(2005\)](#) is quasi-analytic and allows for the calculation of the potential and gravity at any point in a body (this is in contrast to the above approach that is applicable only to radii greater than the maximum radius). This method starts by mapping irregularly shaped density interfaces to spherical ones, and then determines the radial derivatives of the potential and gravitational field on this surface. Using exact values for the potential and gravity field on an interface exterior to the planet (as obtained from a method similar to eqn [30]), these fields are then propagated downwards using a first-order Taylor series approximation. This technique is useful for lithospheric flexure calculations as the net lithospheric load is a function of the potential at the major density interfaces.

A second method for calculating the gravitational field is based upon approximating the shape of a celestial body by a polyhedron. Exact expressions for the potential of a homogeneous polyhedron have been derived by [Werner and Scheeres \(1997\)](#), and expressions for the corresponding spherical harmonic coefficients are given in [Werner \(1997\)](#). The benefit of using this approach is that the resolution of the model (i.e., the spacing between vertices) can be varied according to the resolution of the gravitational field. An application of this method for determining the interior density of an asteroid is given by [Scheeres et al. \(2000\)](#). Alternatively, one could transform eqn [14] into a surface integral by use of Gauss' law and perform the integral numerically for a given shape model (e.g., [Cheng et al., 2002](#)).

10.05.5 Crustal Thickness Modeling

It is well known that the modeling of potential fields is nonunique. For our case, eqns [16] and [32] show that any external gravitational field can be interpreted as a surface density ρb placed at an arbitrary radius D . Nevertheless, by using simplifying assumptions based on geologic expectations, it becomes possible to uniquely invert for parameters related to the interior structure of a planet.

Perhaps the simplest example of such an investigation is the construction of a global crustal thickness model. In this case, the nonuniqueness associated with potential modeling is removed by assuming that the observed gravitational field arises only from relief along the surface and crust–mantle interface (i.e., the ‘Moho’), and that the density of the crust and mantle are constant. It is furthermore required to assume a mean crustal thickness, or to anchor the inverted crustal thickness to a given value at a specific locale. If lateral density variations in either layer could be constrained by other means, then these could easily be incorporated into the model.

The first step is to calculate the Bouguer correction, which is the contribution to the potential of surface relief referenced to the mean planetary radius. Subtracting this from the observed gravitational field yields the Bouguer anomaly, and this is then ascribed to being caused by relief along the crust–mantle interface. To first order, this relief could be determined in the spectral domain by downward continuing the Bouguer anomaly coefficients C_{lm}^{BA} to a radius D , and then setting these equal to those predicted from the mass-sheet approximation of eqn [32]. However, two additional factors generally need to be taken into account in such an analysis. First, downward continuing the Bouguer anomaly amplifies short-wavelength noise that is often present in the observed gravitational field. Second, the first-order mass-sheet approximation may not be sufficiently accurate if the Moho undulations are large.

By minimizing the difference between the observed and predicted Bouguer anomalies, as well as an additional constraint such as the amplitude of the Moho undulations in the spectral domain, the Moho relief can be computed via the following equation (see [Wieczorek and Phillips, 1998](#)):

$$b_{lm} = \omega_l \left[\frac{C_{lm}^{BA} M(2l+1)}{4\pi \Delta\rho D^2} \left(\frac{R}{D} \right)^l - D \sum_{n=2}^{l+3} \frac{(b^n)_{lm} \prod_{j=1}^n (l+4-j)}{D^n n! (l+3)} \right] \quad [33]$$

where $\Delta\rho$ is the density jump across the crust–mantle interface, R is the reference radius of the Bouguer anomaly coefficients, and ω_l is a filter that stabilizes the downward continuation procedure. The filter ω_l should be close to unity for small l and decrease in magnitude with increasing l . While there is no simple

analytic solution to this equation, the relief along the crust–mantle interface can be determined using an iterative approach: first the coefficients b_{lm} are approximated by ignoring the higher-order terms on the right-hand side, then, using this estimate, the higher-order terms are calculated, and a new estimate of b_{lm} is obtained. Examples of crustal thickness models that were obtained using this procedure are shown in **Figure 10** for the Moon, Mars and Venus, and the major modeling assumptions specific to each body are described below. It is important to note that these models do not assume that the crust is isostatically compensated; such a hypothesis could be tested for a given model.

For the Moon, it is known that the mare basaltic lava flows are considerably denser than upper crustal materials (~ 3100 vs 2800 kg m^{-3}), and that these can reach thicknesses of several kilometers within some of the largest impact basins. Thus, when computing the Bouguer anomaly for the Moon, the gravitational attraction of these must be estimated. As a result of the variable spatial resolution of the lunar gravity field, it is also necessary to apply a strong downward continuation filter (see Wieczorek and Phillips, 1998) in order to suppress unphysical Moho undulations that arise on the farside. After truncating the potential and topography coefficients beyond degree 85, and assuming an average crustal thickness of 45 km and a mantle density of 3320 kg m^{-3} , the Moho relief was iteratively determined using eqn [33]. The obtained crustal thickness model displayed in **Figure 10** demonstrates that the thickness of the lunar crust could vary from approximately zero beneath some basins to more than 100 km in the highlands (see Wieczorek *et al.*, 2006). Neglecting the finite amplitude terms in eqn [33] could give rise to errors as large as 20 km (Neumann *et al.*, 1996).

Crustal thickness models for Mars and Venus are also presented in **Figure 10**. The model for Mars is an updated version from Neumann *et al.* (2004) that uses the recent JGM95J01 gravity model. For this model, the low density of the polar caps, the higher than typical densities of the Tharsis volcanoes, and the gravitational attraction resulting from the core flattening were explicitly taken into account. A mean crustal thickness of 45 km was assumed, and in downward continuing the Bouguer anomaly, a filter was constructed such that the power spectrum of the Moho relief resembled that of the surface relief. For the Venusian model, a mean crustal thickness of 35 km was assumed, the potential and topography

coefficients were truncated beyond degree 60, and densities of 2900 and 3330 kg m^{-3} were used for the crust and mantle, respectively. The inclusion of finite amplitude corrections for Venus only affects the obtained crustal thicknesses by a few kilometers.

10.05.6 Admittance Modeling

In the crustal thickness modeling presented above, the nonuniqueness associated with potential modeling was removed by making certain assumptions about the mean crustal thickness and the density of the crust and mantle. These and other parameters can be estimated if one instead assumes that surface topography is supported by a specific mechanism, such as Airy compensation or lithospheric flexure. Using such a model, the relationship between gravity and topography can be determined, and by comparing to the observed values, model parameters can be estimated. As is described in the following two subsections, two methods are in common use; one is based upon modeling the GTR in the space domain, whereas the other models the admittance and correlation functions in the spectral domain.

10.05.6.1 Spatial Domain

One method that has proven to be fruitful for estimating the mean crustal thickness of a planet is modeling of the GTR in the space domain. This technique was initially developed by Ockendon and Turcotte (1977) and Haxby and Turcotte (1978) for the Earth where it was shown that the isostatic geoid anomaly was approximately equal to the vertical dipole moment of density variations within the lithosphere. For the specific cases of Airy and Pratt isostasy, the ratio between the geoid and topography was found to be proportional to the crustal thickness. This method was derived using a Cartesian geometry, and is strictly valid in the long-wavelength limit.

An alternative approach has been developed in spherical coordinates by Wieczorek and Phillips (1997) where it has been shown that the GTR can be approximated by the expression

$$\text{GTR} = R \sum_{l=l_{\min}}^{l_{\max}} W_l Q_l \quad [34]$$

where l_{\min} and l_{\max} correspond to the minimum and maximum spherical harmonic degrees that are

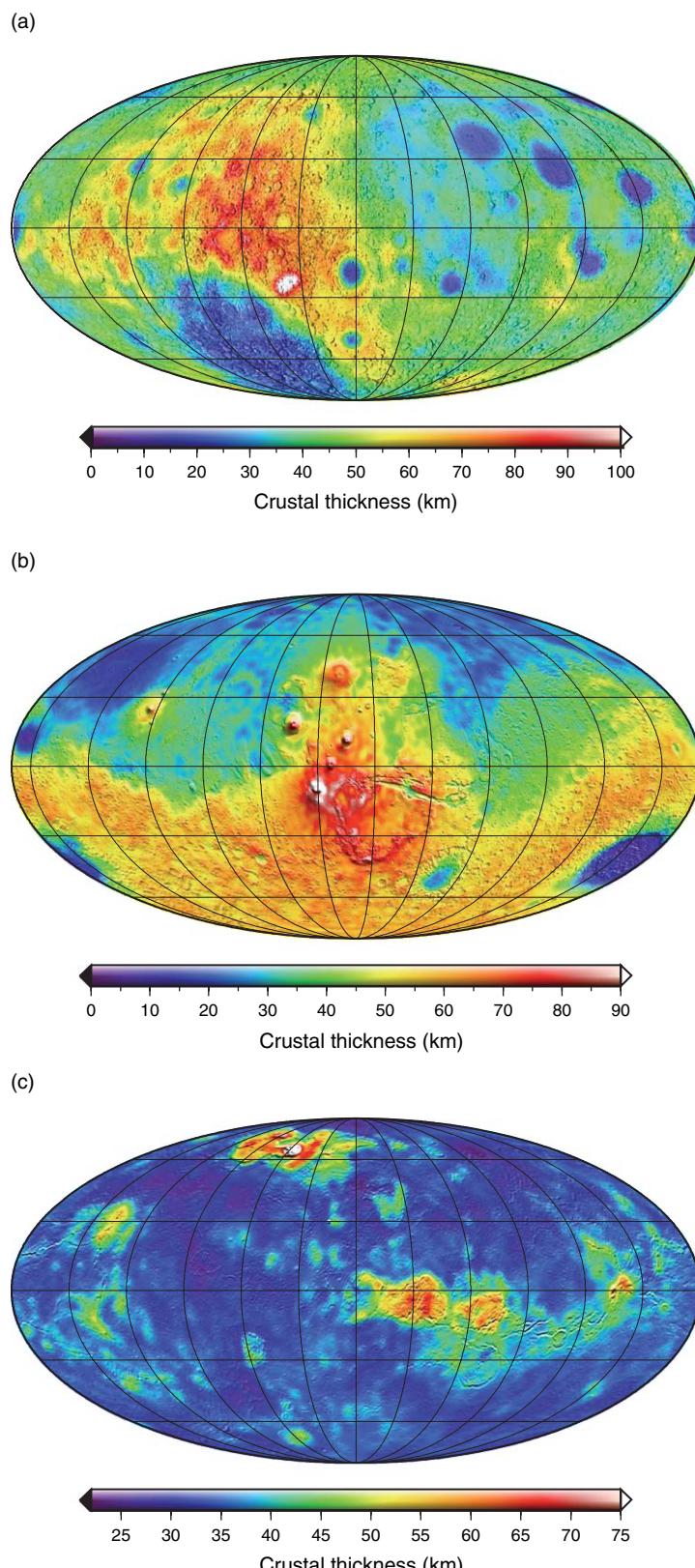


Figure 10 Crustal thickness models for (a) the Moon, (b) Mars, and (c) Venus. See Sections 10.05.5 and 10.05.8 for details.

considered, R is the mean planetary radius, Q_l is the linear transfer function between the potential and topographic coefficients, and W_l is a weighting function that is proportional to the topographic power of degree l ,

$$W_l = S_{bb}(l) / \sum_{i=l_{\min}}^{l_{\max}} S_{bb}(i) \quad [35]$$

The underlying assumption of this model is that the GTR is independent of position for a given compensation model, and this has been empirically validated for the highlands of the Moon and Mars for the case of Airy isostasy (Wieczorek and Phillips, 1997; Wieczorek and Zuber, 2004).

As the power spectra of planetary topography are ‘red’ (i.e., they possess more power at long wavelengths than short wavelengths), eqn [35] shows that the largest contribution to the GTR will inevitably come from the lowest degrees. As an example, approximately 80% of the GTR for the Moon is determined by degrees less than 30. As the topography of the ancient highland crust of a planet is likely to be isostatically compensated at these wavelengths, it is common to employ a model based on the condition of Airy isostasy for these regions. Assuming that the density of the crust is constant, and using the condition of equal mass in crust-mantle columns, it is straightforward to show using eqn [32] that the transfer function between the potential and topographic coefficients is

$$Q_l = \frac{C_{lm}}{b_{lm}} = \frac{4\pi \rho_c R^2}{M(2l+1)} \left[1 - \left(\frac{R-T_c}{R} \right)^l \right] \quad [36]$$

where ρ_c is the density of the crust, and T_c is its mean thickness.

In practice, the GTR is determined by fitting the observations to a straight line within a region that is believed to be consistent with the employed model. By utilizing a plot of the predicted GTR versus T_c (obtained from eqns [34]–[36]), the crustal thickness can then be estimated for a given value of ρ_c . Nevertheless, as the GTR is heavily influenced by the longest wavelength components of the gravity and topography, several aspects need to be carefully considered when performing such an analysis.

It is first necessary to ensure that the entire signal of the geoid and topography are governed by the same compensation model. While this can never be entirely satisfied, certain anomalous long-wavelength features can sometimes be identified and removed. For instance, most planets and satellites possess

significant rotational and/or tidal contributions to their degree-2 shape, and these signatures can be minimized by setting these coefficients to zero. For Mars, in addition to the degree-2 rotational signature, the longest wavelength components have been affected by the load and flexural response associated with the Tharsis province (see Zuber and Smith, 1997; Phillips *et al.*, 2001; Wieczorek and Zuber, 2004). Furthermore, as the degree-1 potential terms are zero when the gravitational field is referenced to the body’s center of mass, any degree-1 topography that exists may need to be treated separately. Finally, as the GTR is largely determined by the longest wavelength components of the geoid and topography, it is necessary that the region of interest be sufficiently large when regressing the geoid and topography data.

10.05.6.2 Spectral Domain

Two major shortcomings associated with modeling the GTR are that only a single wavelength-independent parameter is used (the GTR) and the observed value could be biased by long-wavelength features that are unrelated to the assumed compensation model (such as lateral density anomalies in the mantle caused by convection). An alternative modeling approach that largely bypasses these concerns is to model the relationship between the gravity and topography for a certain region in the spectral domain. As wavelength-dependent admittance and correlation functions are obtained, in principle, it is possible to invert for several model parameters. The major shortcoming is that the resolution of the gravity and topography becomes increasingly poor with increasing degree l . This section describes the basic concepts involved with such a spectral analysis applied on a global scale. In Section 10.05.7, the technique of obtaining localized spectral estimates will be described.

Let us presume that the potential and topography coefficients are related via an equation of the form

$$C_{lm} = Q_{lm} b_{lm} + I_{lm} \quad [37]$$

where Q_{lm} is a linear nonisotropic transfer function, and I_{lm} is that portion of the potential not described by the model. For simplicity, it will be assumed that I_{lm} is zero (though this assumption will be relaxed later) and that the topography is noise free. Though the above relationship is inherently nonisotropic, it is often useful to work with the power

and cross-power spectra of the gravitational field and topography, S_{bb} , S_{gg} and S_{bg} , which only depend upon spherical harmonic degree l . The goal is to fit these three functions to those obtained from an appropriate model. In order to remove the model dependence of certain nuisance parameters, it is convenient to work with ratios of these three cross-power spectra. Although several such ratios involving powers of these are possible, only two will be independent, and it is traditional to use the admittance and correlation spectra as previously defined by eqns [25] and [26]. If a model describing a planet's gravity and topography is to be considered successful, then it must satisfy both of these functions. If one or both of these functions cannot be fit for a given degree, then this is a clear indication that either the model assumptions are inappropriate for the region being investigated, or the input data sets are not sufficiently accurate.

If one treats the lithosphere of a planet as a thin elastic spherical shell overlying a fluid interior (see Kraus, 1967), a simple relationship exists in the spectral domain between the load on the lithosphere and its deflection (see Turcotte *et al.*, 1981; Willemann and Turcotte, 1981; Banerdt, 1986). If loading at only a single interface is considered (either at or below the surface), then the transfer function in eqn [37] is isotropic (i.e., independent of m). For this situation, expressions for the admittance and correlation functions can be schematically written as:

$$Z(l) = f(\rho_c, \rho_m, \nu, E, T_e, T_c, z, g, R) \quad [38]$$

$$\gamma(l) = 1 \quad \text{or} \quad -1 \quad [39]$$

where f denotes a functional dependence on the enclosed parameters. In particular, the admittance function explicitly depends on the crustal and mantle density, Poisson's ratio ν , Young's modulus E , the elastic thickness T_e , the crustal thickness, the depth of the load z , the magnitude of the gravitational acceleration g , and the radius of the planet. For an isotropic transfer function Q_l , it is trivial to show that the degree-correlation function (in the absence of noise) is equal to the sign of Q_l . This model has been amended by McGovern *et al.* (2002) and Belleguic *et al.* (2005) to include inphase loads applied to and below the surface when the two are linearly related by a degree-independent constant. Such models would include an additional parameter L , which is a function of the relative magnitudes of the surface and subsurface loads. Geologic situations where surface and subsurface loads might be

perfectly correlated include isolated volcanoes and impact basins.

An alternative loading model that includes loads applied to and below the surface was developed by Forsyth (1985) in the Cartesian domain (see also Banks *et al.*, 2001). In contrast to models that take the applied loads to be perfectly in phase, he assumed that the phase differences of the applied surface and subsurface loads were random. Such an assumption might be expected to be reasonable for continental cratons where several geologic processes have operated over extended periods of time (such as erosion, sedimentation, and magmatism). In contrast to eqn [39], this model possesses a wavelength-dependent correlation function. A model similar to that of Forsyth (1985) has been derived in spherical coordinates by Wieczorek (submitted manuscript), and can be schematically described by the following equations:

$$Z(l) = f(\rho_c, \rho_m, \nu, E, T_e, T_c, z, L, \alpha_l, g, R) \quad [40]$$

$$\gamma(l) = f'(\rho_c, \rho_m, \nu, E, T_e, T_c, z, L, \alpha_l, g, R) \quad [41]$$

where both f and f' represent generic functional dependencies. As an extension to Forsyth's model, this formulation allows for an arbitrary phase relationship between the applied surface and subsurface loads that is described by the additional parameter α . The expectation of this function is given by the expression

$$\alpha_l = \frac{\sum_{m=-l}^l \langle \cos \Delta_{lm} \rangle}{(2l+1)} \quad [42]$$

where Δ_{lm} denotes the phase difference between the two loads, and $\langle \dots \rangle$ is the expectation operator. For random phases, α is zero, and the model degenerates to that of Forsyth (1985). When the loads are perfectly in or out of phase by 0 or 180° , $\alpha = \pm 1$ and the model is analogous to that of McGovern *et al.* (2002) and Belleguic *et al.* (2005).

Examples of the predicted free-air admittance and correlation functions are shown in Figure 11 for several values of the elastic thickness and phase parameter α . These models were generated using parameters typical for the planet Mars, and the magnitudes of the applied surface and subsurface loads were chosen to be equal. As is seen, these curves are strong functions of both the elastic thickness and α , and by considering both the admittance and correlation, it may be possible to separate the effects of the two. The free-air correlation function is seen to possess low values over a restrictive range of

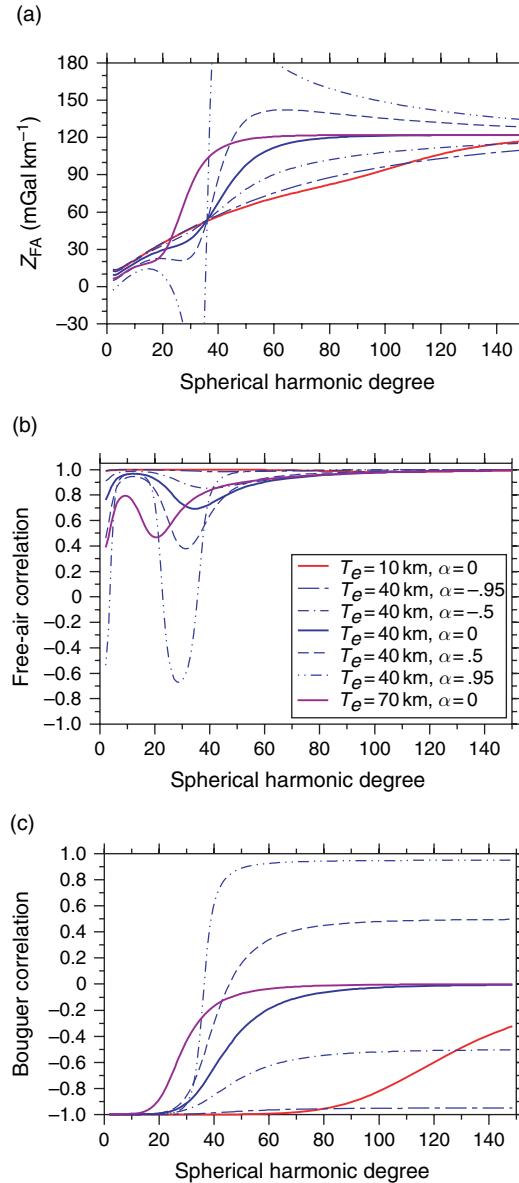


Figure 11 (a) Free-air admittance, (b) free-air correlation, and (c) Bouguer correlation, for a flexural model with equal magnitudes of applied surface and subsurface loads. Model parameters correspond to the planet Mars, with $T_c = z = 50$ km, $\rho_c = 2900$ kg m $^{-3}$, $\rho_m = 3500$ kg m $^{-3}$, and $E = 10^{11}$ Pa. Solid lines correspond to the case where the applied surface and subsurface loads have random phases (i.e., $\alpha = 0$), and the dashed lines correspond to the case where these loads are partially correlated.

wavelengths that is diagnostic of the elastic thickness. Furthermore, the free-air correlation function is seen to approach unity at large degrees ($l \gtrsim 100$). In practice, if a decline of the free-air correlation is observed with increasing degree, this is usually a

good indicator of a loss of fidelity with the employed gravitational model. The predicted Bouguer correlation function is also shown for the same model parameters, and this shows a behavior similar to that predicted by Forsyth's model. In particular, it is noted that the Bouguer correlation asymptotically approaches the value of α with increasing degree. While the Bouguer correlation is useful for interpretive purposes, its use is not advocated here because the Bouguer gravitational anomaly critically depends upon the value chosen for the crustal density, and this is generally not known *a priori*.

The preceding discussion explicitly assumed that the unmodeled gravitational signal I_{lm} in eqn [37] was identically zero. As mentioned in Section 10.05.2.2, if I_{lm} and the topography are uncorrelated, the expectation of the admittance will not be affected by the presence of such a signal. However, the expectation of the gravitational power spectrum will be biased upwards by an additive constant S_H , and this will tend to bias the correlation function downwards. Fortunately, any theoretical loading model can be easily modified to incorporate an unmodeled signal I_{lm} that is uncorrelated with the topography: all one needs to do is add a degree-dependent constant S_H to the theoretically predicted gravitational power spectrum (see eqn [24]). It would be a simple matter to include S_H as an inversion parameter in a generic loading model such as eqn [41], though it is noted that this has not yet been attempted. While I_{lm} is commonly assumed to be gravitational measurement noise, this need not be the case. As described by McKenzie (2003), massive erosion or sedimentation of a lithosphere possessing surface and subsurface loads could give rise to a final state where the surface is perfectly flat, but where gravitational anomalies are present. If this final state were taken as the initial condition of a subsequent loading event, then the initial gravitational signature (which is unrelated to the second loading model) would be expected to be uncorrelated with the subsequently generated topography.

Finally, it is important to reemphasize that if a given model of lithospheric loading is an accurate description of reality, it must fit both the admittance and coherence functions. If this can not be done, then either the model or data must be inaccurate. Unfortunately, the vast majority of published investigations that use Forsyth-like loading models employ only the admittance or coherence function. The values of inverted parameters from such studies, while perhaps correct, need confirmation by analysis of the other function. Notable exceptions include

the papers by Forsyth and coworkers (Forsyth, 1985; Bechtel *et al.*, 1987, 1990; Ebinger *et al.*, 1989; Zuber *et al.*, 1989; Phillips, 1994; Pérez-Gussinyé *et al.*, 2004). Similarly, many published investigations that employ a loading model with only surface or subsurface loads also ignore the correlation function, even though such models explicitly require this to be ± 1 when I_{lm} is zero. Nevertheless, it must be noted that no published study has yet attempted to fit both the admittance and correlation functions simultaneously when an unmodeled gravitational signal is present that is uncorrelated with the topography.

10.05.7 Localized Spectral Analysis

As the spherical harmonics are global basis functions, the power spectrum as defined by eqn [9] is necessarily representative of the global properties of the function. In geophysics, however, it is reasonable to suspect that the spectral properties of the gravity and topography will vary as a function of position. For example, the elastic thickness may differ among geographic provinces as a result of their unique histories. Alternatively, it might arise that the data are only locally known, and that one would like to estimate the power spectrum based exclusively upon these data.

One way of obtaining localized estimates of a function's power spectrum is to first multiply the data by a localizing window (commonly referred to as a data taper), and then to expand this windowed function in spherical harmonics (for a detailed discussion in the Cartesian domain, see Percival and Walden (1993)). However, as a result of the windowing procedure, the resultant power spectrum will differ from the true value. For the case where the input field is stationary, and the spherical harmonic coefficients are governed by a zero-mean stochastic process, it can be shown that the expectation of the windowed power spectrum is related to the global spectrum by the following relation (Wieczorek and Simons, 2005, 2007):

$$\langle S_{\Phi\Gamma}(l) \rangle = \sum_{j=0}^L S_{bbj} \sum_{i=|l-j|}^{l+j} S_{fg}(i) \left(C_{ij0j}^{lo} \right)^2 \quad [43]$$

Here, b represents an arbitrary window bandlimited to degree L , Φ and Γ are the windowed fields bf and bg respectively, and the symbol in parentheses is a Clebsch-Gordan coefficient (these are related to the Wigner 3- j symbols and are proportional to the

integral of three Legendre functions, see Varshalovich *et al.* (1988)). The expectation of the windowed power spectrum $S_{\Phi\Gamma}$ is seen to be related to the global spectrum by a smoothing operation reminiscent of a convolution.

For a localized spectral analysis, the question naturally arises as to what form the localizing window should take. In order to localize the data, it is clear that the amplitude of the window (or its power) should be near zero outside the region of interest. Furthermore, as a result of the convolution-like relationship between the global and windowed spectra, the bandwidth L of the window should be as small as possible in order to limit this spectral smoothing. Slepian and coworkers (see Slepian, 1983) previously posed and solved this problem in Cartesian geometry by finding those windows whose power were optimally concentrated in a specified region. Using this same criterion, Wieczorek and Simons (2005) and Simons *et al.* (2006) solved for those bandlimited windows on the sphere that are optimally concentrated within a spherical cap of angular radius θ_0 . This optimization problem reduces to a simple eigenvalue equation whose solution yields a family of orthogonal data tapers; the quality of the concentration is given by the corresponding eigenvalue. For the case of zonal functions, it was shown that the properties of these windows depend almost exclusively on the space-bandwidth product

$$N_0 = (L+1) \frac{\theta_0}{\pi} \quad [44]$$

and that the first $N_0 - 1$ windows were near optimally concentrated. As an example, the best three concentrated windows corresponding to $\theta_0 = 40^\circ$, $N_0 = 4$, and $L = 17$ are plotted in Figure 12. The number of well concentrated windowing functions can be dramatically increased by making use of the nonzonal tapers (Wieczorek and Simons, 2007). The extension of this method to arbitrarily-shaped concentration regions is given by Simons *et al.* (2006).

While a method had previously been used in spherical coordinates where a function is multiplied by a single localization window (Simons *et al.*, 1997), the use of multiple tapers, as originally pioneered by Thomson (1982) in the Cartesian domain, has several key advantages. First, while the energy of any single window will nonuniformly cover the concentration region, the cumulative energy of orthogonal tapers is nearly constant for the region of interest. Thus, an average of spectra obtained from several orthogonal tapers will be more representative of the data than

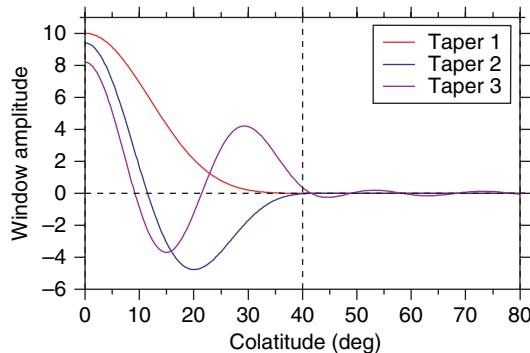


Figure 12 The first three zonal data tapers, bandlimited at $L = 17$, whose power is optimally concentrated within colatitudes less than 40° . The space-bandwidth product N_0 is here equal to 4.

that of a single taper. Second, while it is usually not possible to obtain the expectation of the localized spectrum since there is generally only a single field available for analysis, the spectral estimates obtained from orthogonal tapers are nearly uncorrelated, and their average can be considered as an approximation of the expectation. Finally, by using multiple tapers, it is possible to make an estimate of the uncertainty associated with a given spectral estimate; this is expected to decrease as approximately $1/\sqrt{N}$ where N is the number of tapers (Wieczorek and Simons, 2005, 2007).

In performing a localized spectral analysis, there are several factors that need to be considered. First, it is noted that if the fields f and g of eqn [43] are only known to a maximum spherical harmonic degree L_{fg} , then only the first $L_{fg} - L$ windowed spectral estimates are reliable. Second, those localized spectral estimates with degrees less than L are heavily biased and possess relatively large uncertainties, making these of little use for geophysical analysis. Third, while a multitaper spectral analysis is generally preferable to using a single taper, the above two concerns present serious limitations when working with the relatively low resolution gravity fields of Venus, Mars and the Moon. Depending on the size of the concentration region, it may be infeasible to use the larger bandwidths that are required for obtaining several well concentrated tapers.

Finally, when comparing model results to observations, it is emphasized that the two must be windowed in the same manner (e.g., Pérez-Gussinyé *et al.*, 2004). If the analysis is performed by generating forward models of the gravity field using

the known topography, then it is only necessary to localize these functions in the same manner as the data. Alternatively, if no explicit expression exists for Q_{lb} (as in the model of Forsyth (1985) and that presented in Section 10.05.6.2, both of which are statistical in nature), then it is necessary to window the predicted (cross-) power spectra using eqn [43] before calculating the theoretical admittance and correlation functions (for the Cartesian analog, see equation 4.2 of Thomson, 1982).

10.05.8 Summary of Major Results

10.05.8.1 Earth

The gravity and topography of the Earth have been used extensively to decipher the rheological properties of the crust and upper mantle. The literature is voluminous, and the reader is referred to several reviews in Volumes 3 and 6 of this series, Watts (2001), and the references in the papers cited below. Here, only a few subjects will be touched upon that bear relevance to investigations of Venus, Mars, and the Moon. These include modeling of the elastic thickness of the oceanic and continental lithosphere, inelastic flexural modeling, and the modeling of dynamic topography and geoid signatures associated with mantle convection.

Flexural modeling of the oceanic lithosphere is relatively simple in that loading is primarily a result of the construction of isolated shield volcanoes. Elastic thickness estimates have been obtained by modeling the topographic and gravity signatures of these features, and it is widely accepted that the elastic thickness is primarily dependent upon the age of the plate at the time of loading, with T_e being generally less than about 45 km (for a review, see Watts, 2001). In particular, a plot of the elastic thickness versus age of the lithosphere at the time of loading resembles the time dependence of the depth to an isotherm ($\sim 300\text{--}600^\circ\text{C}$) predicted from a plate cooling model (see Watts and Zhong, 2000). This suggests that the flexural signature has been ‘frozen’ into the lithosphere as it cooled and that long term viscoelastic relaxation has been relatively unimportant. Nevertheless, a description of the initial short-term subsidence of the lithosphere (i.e., the first few 10s of ka) requires the use of a viscoelastic model, and given the relatively restricted age range of oceanic lithosphere (<200 Ma), it is difficult to discern if viscoelastic relaxation would be important at longer timescales. It is important to note that most flexural

modeling of features on the other terrestrial planets has been for loads that were emplaced on the lithosphere over a billion years ago.

Investigations of the continental elastic thickness have been more contentious. Part of the difficulty arises because it is not clear *a priori* as to the importance of subsurface loading and the phase relationship of the surface and subsurface loads (see Section 10.05.6.2). A loading model was developed by Forsyth (1985) that took into account both surface and subsurface loading under the assumption that the two were uncorrelated, and application of this method has yielded elastic thicknesses in the broad range of 5–134 km (Forsyth, 1985; Bechtel *et al.*, 1987, 1990; Ebinger *et al.*, 1989; Zuber *et al.*, 1989; Pérez-Gussinyé *et al.*, 2004). There is currently some controversy as to whether the values greater than ~ 25 km are reliable (compare McKenzie (2003) with Watts and Burov (2003)), but this debate will not be definitively settled until investigators fit both the admittance and correlation functions simultaneously while taking into account the presence of an unmodeled gravitational signal that is uncorrelated with the topography (see Section 10.05.6.2). Furthermore, some studies that have inverted for the elastic thickness using multitaper spectral analysis techniques have done so using a methodology that is not entirely correct. In particular, the windowed power spectra from a multitaper analysis represent a convolution of the true power spectra with that of the window (see equation 4.2 of Thomson, 1982). Thus, it is necessary to convolve the theoretical (cross-) power spectra with that of the data tapers before obtaining theoretical windowed admittance and coherence functions. Regardless, application of the multitaper method has convincingly shown that the elastic thickness of some continental regions is not always isostropic (e.g., Simons *et al.*, 2000, 2003), which is an assumption common to most studies.

While the majority of investigations that model flexure of the lithosphere assume that it is perfectly elastic, elastic stresses are often predicted to be in excess of the strength of geologic materials. A simple modification to the elastic flexure equation that takes this into account is to replace the elastic bending moment-curvature relationship with one that is based upon an elastic-perfectly plastic (EP) model of the lithosphere's yield strength (e.g., Burov and Diament, 1995; Mueller and Phillips, 1995). In particular, the strength of the upper crust is limited by brittle failure, and stresses in the lower crust and mantle are limited by their ductile strength for a

specified strain rate. Predicted flexural profiles are time-invariant and can sometimes differ significantly from those of the perfectly elastic model. As the ductile strength is temperature dependent, these results are sensitive to the assumed lithospheric temperature gradient.

A more realistic model of lithospheric deformation uses a time-dependent elastoviscoplastic (EVP) formulation (e.g., Albert and Phillips, 2000; Albert *et al.*, 2000; Brown and Phillips, 2000). The main advantage of these models is that the strain rates are explicitly calculated, as opposed to assumed as in the EP models. While the best-fit EP and EVP flexural profiles can be quite similar, it is not clear *a priori* how one should estimate the characteristic strain rate that is required for the EP model without running a full EVP simulation (Albert *et al.*, 2000). The EVP models show that significant decoupling of stresses may occur between the crust and mantle if the lower crust is sufficiently weak (e.g., Brown and Phillips, 2000). When this occurs, the effective elastic thickness decreases; the exact value is highly dependent upon the crustal thickness, load magnitude, and assumed rheology of the crust and mantle. In contrast, when the lower crust is strong, the maximum achievable effective elastic thicknesses are consistent with the depth of an $\sim 700^\circ\text{C}$ isotherm obtained from a lithospheric cooling model. Flexural modeling of a volcano growing on a cooling lithosphere shows that the effective elastic thickness is 'frozen' into the lithosphere shortly after volcanic construction is complete (Albert and Phillips, 2000).

Finally, in addition to near-surface crustal thickness and density variations, significant gravity and topography signatures can be generated by dynamic processes in the mantle, such as beneath hot spots and subduction zones. While there are few, if any, convincing examples of plate subduction on the other terrestrial planets, hot spots similar to the Earth are believed to exist on both Venus and Mars. Dynamic modeling of plumes shows that the major variable controlling the surface gravity and topography signatures is the depth dependence of the mantle viscosity. In the absence of a shallow low viscosity asthenosphere, convective stresses generated at depth are efficiently coupled to the surface, generating large signals and large corresponding effective depths of compensation. However, the inclusion of a shallow low viscosity zone can significantly reduce these signatures, and apparent depths of compensation (ADCs) are found to be significantly shallower (e.g., Robinson and Parsons, 1988; Ceuleneer *et al.*,

1988). Joint inversions utilizing mantle density anomalies from seismic tomography and estimates for the dynamic topography signal imply the existence of a low-viscosity zone somewhere in the upper mantle, and a gradual increase in viscosity with depth by an order of magnitude in the lower mantle (e.g., Panasyuk and Hager, 2000).

10.05.8.2 Venus

Our knowledge of Venus has dramatically improved since the acquisition of gravity, topography, and SAR imagery by the Magellan mission between 1990 and 1994. While the size and bulk density of Venus were known beforehand to be similar to that of the Earth, this planet was found to differ dramatically in that it lacks any clear sign of plate tectonics. A major unanswered question is how this planet loses its internal heat, and whether or not this process is episodic or uniform in time. Geophysical analyses have been used to constrain the crustal and elastic thickness, and the latter has been used to place constraints on the temperature gradient within the lithosphere. Reviews concerning the geophysics of this planet can be found in Phillips *et al.* (1997), Grimm and Hess (1997), Schubert *et al.* (1997), and Nimmo and McKenzie (1998).

The crustal plateaus of Venus generally have low-amplitude gravitational and topographic signatures within their interiors and are potential candidates for being isostatically compensated (one notable exception is Ishtar Terra). By assuming that the surface topography is compensated at a single interface, Smrekar and Phillips (1991) obtained best-fit ADCs between 50 and 90 km by modeling Pioneer Venus line-of-sight gravity data over Gula, Ovda, Phoebe, Tellus, Thetis and Ulfrun Regiones. Using higher resolution Magellan data Grimm (1994) obtained best-fit ADCs between 20 and 50 km for Alpha, Tellus, Ovda, and Thetis Regiones. These values are plausibly interpreted as representing the crust–mantle interface, especially when considering that the crustal thickness at the mean planetary radius would be thinner given the high average elevations associated with the above study regions. An analysis of GTRs by Kucinskas and Turcotte (1994) found zero-elevation crustal thickness of 50 ± 7 and 65 ± 13 km for the crustal plateaus of Ovda and Thetis Regiones, respectively, consistent with the above mentioned studies. Moore and Schubert (1997) similarly found values 44 ± 4 , 83 ± 8 , 75 ± 17 , and 75 ± 22 km for Alpha, Ovda, Tellus, and Thetis Regiones,

respectively. (These latter estimates might overestimate the zero-elevation crustal thickness as regional low-pass filtered versions of the gravity and topography were removed before performing the regression.) If any portion of the geoid and topography were to result from Pratt or thermal compensation, the obtained crustal thicknesses would represent an upper bound. A spectral admittance study by Phillips (1994) (described below) obtained a slightly thinner crustal thickness of 30 ± 13 km for the region of Atla Regio (updated to 38 ± 9 km in Phillips *et al.* (1997)).

An additional constraint concerning the thickness of the Venusian crust is related to its compositional buoyancy. In particular, if the crust were basaltic in composition, then this material should undergo a phase transition at high pressure to the more dense mineral assemblage of eclogite (e.g., Grimm and Hess, 1997). This material could potentially delaminate from the crust as a result of its high density, and the depth of this phase transition might thus constrain the maximum achievable crustal thickness. For a MORB composition, the eclogite phase transition is predicted to occur at depths of ~ 70 to 120 km (see Ghent *et al.*, 2004) for linear temperature gradients of 5 and 15 K km^{-1} , respectively. Inspection of the crustal thickness map in Figure 10 (which is based upon the premise of an average crustal thickness of 35 km) shows that crustal thicknesses near 70 km exist only in the highland plateaus of Ishtar Terra, and Ovda and Thetis Regiones, suggesting that crustal delamination could have occurred in these regions. If this crustal thickness map is an accurate representation of reality, it is curious that the highland crust is almost exactly twice as thick as the low-land crust; this potentially could arise by the obduction of one plate over another.

In contrast to the majority of the highland plateaus, large ADCs and GTRs have been found for the volcanic rises (Smrekar, 1994; Kucinskas and Turcotte, 1994; Moore and Schubert, 1995, 1997) and Ishtar Terra (Grimm and Phillips, 1991; Hansen and Phillips, 1995; Moore and Schubert, 1997). Such values are not consistent with compensation occurring solely by crustal thickening, but require some form of dynamic support from the mantle via stresses induced by ascending mantle plumes (e.g., Vezolainen *et al.*, 2004), and/or substantial thinning of a thick (~ 300 km) thermal lithosphere (Kucinskas and Turcotte, 1994; Moore and Schubert, 1995, 1997). If a low viscosity asthenosphere were present at shallow mantle depths, as is the case of the Earth, the predicted GTRs and ADCs resulting

from dynamic support would be considerably smaller than measured as a result of the decoupling of stress between the lithosphere and mantle (e.g., Kiefer *et al.*, 1986; Kiefer and Hager, 1991). These results seem to imply that in contrast to the Earth, Venus lacks a low viscosity zone, which is most likely a result of a volatile-poor mantle. Modeling by Pauer *et al.* (2006) suggests that the viscosity of the Venusian mantle could increase by a factor of about 10–80 from the upper to lower mantle, similar to that of the Earth. A strong coupling of stresses between the lithosphere and mantle is the likely cause of the high spectral correlation between gravity and topography for the lowest spherical harmonic degrees of Venus (contrast **Figures 2 and 4**).

Elastic thickness estimates have been obtained for a variety of features based exclusively on topographic profiles that are indicative of flexure. The benefit of using topography alone is that small features can be investigated that are not resolved in the current gravity model. Elastic thicknesses of 11–25 km have been obtained by modeling the Freyja Montes foredeep (Solomon and Head, 1990; Sandwell and Schubert, 1992), and ~10–60 km for potential flexural bulges outboard of coronae (Sandwell and Schubert, 1992). Additional features modeled by Johnson and Sandwell (1994) yield elastic thicknesses of 10–40 km, and potential subduction related sites possess a range of 6–45 km (Schubert and Sandwell, 1995). Predicted stresses are largest where the plate curvature is greatest, and faulting is generally visible in the Magellan SAR imagery at these locations. Modeling by Barnett *et al.* (2002) yielded best-fit elastic thicknesses that are consistent with the above studies. Modeling the location of concentric faulting around Nyx Mons (a volcano in Bell Regio) implies a best-fit elastic thickness of ~50 km (Rogers and Zuber, 1998). It is important to note that the assumption of a perfectly elastic rheology may be grossly inappropriate for some features. For instance, the magnitude of the flexure-induced bulge south of Artemis Chasma implies that significant yielding has occurred within the lithosphere, and inelastic modeling by Brown and Grimm (1996) shows that a significant compressive in-plane force is required at this locale.

The elastic thickness has also been estimated for various regions of Venus through a combined analysis of gravity and topography data in the spectral domain. Unfortunately, with the exception of one study, these investigations have only fit the admittance or coherence functions, but not both simultaneously (e.g., McKenzie, 1994; Smrekar, 1994; Simons *et al.*, 1994,

1997; McKenzie and Nimmo, 1997; Smrekar and Stofan, 1999; Barnett *et al.*, 2000, 2002; Lawrence and Phillips, 2003; Smrekar *et al.*, 2003; Hoogenboom *et al.*, 2004, 2005). While the conclusions of these investigations may be correct, the robustness of the inverted parameter values, as well as the fidelity of the gravity model as a function of wavelength, is difficult to assess. The exception is that of Phillips (1994) who investigated the lithospheric properties of Atla Regio, which is believed to be an active hotspot based on its geomorphology and previously determined large ADCs. Using the loading model of Forsyth (1985), which assumes uncorrelated surface and subsurface loads, it was shown that a single mode of compensation could not explain the entire wavelength range of the admittance and coherence functions. An inversion utilizing only the short wavelengths yielded a crustal thickness of 30 ± 13 km and an elastic thickness of 45 ± 3 km. While surface loading by the volcanic constructs in this area dominates, about 10% of the load is required to be located at shallow depths within the crust. For the long-wavelength range, only the depth of the subsurface load was well constrained with a value near 150 km, and the elastic thickness was constrained only to be less than 140 km.

Finally, by using the obtained elastic thicknesses, or by forward modeling of inelastic flexure, it is possible to place constraints on the crustal temperature gradient at the time of loading. The basic approach is to match the bending moment implied by the elastic model to that predicted by an inelastic rheology (McNutt, 1984). While the obtained temperature gradient estimates lie in the rather broad range of $3\text{--}26\text{ K km}^{-1}$ (Sandwell and Schubert, 1992; Johnson and Sandwell, 1994; Phillips, 1994; Brown and Grimm, 1996; Phillips *et al.*, 1997), the majority of these lie on the low end, between ~4 and 10 K km^{-1} . This is considerably lower than the expected Earth-scaled temperature gradient of $\sim 15\text{ K km}^{-1}$ (e.g., Phillips, 1994), especially when considering that some of these estimates were obtained where an underlying mantle plume might be expected. While such calculations are critically dependent on the validity of the inelastic strength model, the assumed strain rate, and the depth dependence of temperature, these results seem to imply that the background heat flow of Venus is much less than would be expected by analogy to the Earth. Such an interpretation is consistent with a model in which the Venusian lithosphere formed catastrophically $\sim 500\text{--}1000\text{ My}$, and has since been conductively cooling (cf. Parmetier and Hess, 1992; Turcotte, 1995; Moresi and Solomatov, 1998).

However, such a model is not required, or even preferred, by the cratering history of the Venusian plains (Hauck *et al.*, 1998).

10.05.8.3 Mars

Following the acquisition of high resolution gravity and topography data from the Mars Global Survey and Mars Odyssey missions, a number of studies have been published bearing on the crustal and lithospheric structure of Mars. These investigations have placed constraints on the thickness of the Martian crust, the crustal density, and the elastic thickness, and also imply the existence of dynamic support of topography and buried mass anomalies. Reviews concerning the gravity, topography and crust of Mars can be found in Esposito *et al.* (1992), Banerdt *et al.* (1992), Zuber (2001), Wieczorek and Zuber (2004), and Nimmo and Tanaka (2005).

The average thickness of the Martian crust has been constrained by the analysis of GTRs over the ancient southern highlands. After removing the long-wavelength flexural and load signatures associated with the Tharsis province, a zero-elevation thickness of 57 ± 24 km was obtained under the assumption of Airy isostasy (Wieczorek and Zuber, 2004). This range of values is consistent with estimates based upon the viscous relaxation of topography (Nimmo and Stevenson, 2001; Zuber *et al.*, 2000a) and geochemical mass-balance arguments, both of which require the crust to be less than ~ 100 km thick (see Wieczorek and Zuber, 2004). Crustal thickness modeling further requires the mean thickness of the crust to be greater than 32 km.

A global model of the crustal thickness of Mars has been constructed by Neumann *et al.* (2004) by assuming a mean thickness of 45 km (see Figure 10). If their assumption of a constant density crust is correct, the crust of the southern highlands is predicted to be thicker by about 30 km than the northern lowlands. However, if the northern lowland crust is denser than the southern highlands, as implied by the results of Belleguic *et al.* (2005), then the actual crustal thickness difference could be significantly less. The Tharsis province is seen to possess a relatively thick crust, indicative of prolonged volcanic construction, whereas the crust beneath the major impact basins is considerably thinned, and in some places nearly absent.

Localized spectral admittance and correlation spectra have been modeled in spherical coordinates for various regions using the techniques of Simons *et al.* (1997) and Wieczorek and Simons (2005). In the

investigations of McGovern *et al.* (2002, 2004) and Belleguic *et al.* (2005), a thin elastic spherical shell loading model was employed that depended upon the shell's elastic thickness, the load density, the crustal density, and the ratio of the magnitudes of subsurface and surface loads, which were assumed to be in or out of phase by 0° or 180° . When the load density differs from that of the crust, the methodology of Belleguic *et al.* (2005) is superior, and their results are here summarized.

Of all the parameters considered by Belleguic *et al.* (2005), the load density of the major Martian volcanoes was found to be the best constrained with a value of $\sim 3200 \pm 100$ kg m $^{-3}$. This range is consistent with density estimates of the Martian meteorites, which are thought to be derived from these regions based on their young ages, after the inclusion of a few percent porosity. Elastic thickness estimates are somewhat variable, but were found to lie between about 50 and 100 km when only surface loads were considered. However, when both surface and subsurface loads were modeled, only upper and lower bounds could be specified for most regions. The crustal density was constrained only beneath the Elysium rise (which is located in the northern lowlands), and was found to be identical to the density of the superposed load. Based on rock compositions at the Mars Pathfinder site, Neumann *et al.* (2004) have suggested that the southern highland crust could possess a density close to 3000 kg m $^{-3}$. If this inference is correct, and if the crustal density beneath the Elysium rise is representative of the northern lowlands, then this implies a hemispheric dichotomy in crustal composition. This result is consistent with geochemical maps obtained from the Mars Odyssey gamma ray spectrometer that show the northern lowlands possessing a higher iron concentration than the southern highlands (Taylor *et al.*, 2006). The low elevations of the northern plains could thus be a (partial) result of Pratt compensation. Finally, this study found that the inclusion of less dense subsurface loads (either compositional or thermal in origin) improved the misfit between the modeled and observed admittance functions.

Elastic thickness estimates for other locales have been determined using a variety of techniques, but these generally contain a larger number of assumptions. Modeling of the geologically inferred flexural moat of the northern polar cap suggests an elastic thickness between 60 and 120 km (Johnson *et al.*, 2000). If the topography of the dichotomy boundary is flexural in origin, with loading in the northern

plains, then elastic thicknesses of \sim 31–36 km are implied (Watters, 2003). Modeling of 1-D Cartesian admittance functions have been performed by McKenzie *et al.* (2002) and Nimmo (2002), but the validity of the loading model was not tested by calculating theoretical correlation functions. The gravity and topography of the major Martian volcanoes were also modeled in the space domain by Arkani-Hamed (2000), but the finite-amplitude corrections of the modeled gravity field were not included.

One distinctive feature of Mars is the large geoid and topography signals associated with the Tharsis province (e.g., Phillips *et al.*, 2001). Two possible end-member explanations for this observation are that it is either a result of voluminous extrusive lavas that are partially supported by the lithosphere, or dynamic topography associated with an underlying plume. Visco-elastic modeling of the geoid and topography response to internal buoyant loads implies that a plume is incapable of producing the totality of the observed signals (Zhong, 2002; Roberts and Zhong, 2004). By modeling the contributions of both surface and plume signals with a method that approximates a visco-elastic response, the degree 2 and 3 GTRs imply that a plume can only account for \sim 15 and 25% of the geoid and topography signals, respectively. Using a modified approach that includes all spherical harmonic degrees, Lowry and Zhong (2003) inverted for the relative contributions of surface and internal loads and found that a plume could only account for a maximum of \sim 25 and 50% of the observed geoid and topography, respectively.

Finally, it is noted that buried mass anomalies can be investigated by examination of the residual gravity field after subtraction of an appropriate reference model. For instance, by modeling the gravity field of the Syrtis Major region by surface loading of an elastic shell, an unmodeled localized density anomaly was found beneath this volcanic province. The amplitude of this unmodeled anomaly is consistent with the presence of dense cumulates of an extinct magma chamber (Kiefer, 2004). Using a similar technique, buried mass anomalies have also been inferred along portions of the dichotomy boundary (Kiefer, 2005).

10.05.8.4 The Moon

The Clementine and Lunar Prospector missions have significantly improved our knowledge of the Moon's gravity and topography. Unfortunately, the

resolution of the gravity field varies dramatically between the near and farside hemispheres, and the topography derived from the Clementine altimeter is the poorest among the Earth, Venus and Mars. Indeed, the resolution of the nearside gravity field exceeds that of the topography model. Most investigations have concentrated on mapping crustal thickness variations and quantifying the attributes of the nearside impact basins and craters. A comprehensive review can be found in Wieczorek *et al.* (2006).

The thickness of the lunar crust has been estimated by the analysis of GTRs over the nearside highland crust (Wieczorek and Phillips, 1997). Assuming that these regions are Airy compensated, updated results presented by Wieczorek *et al.* (2006) imply an average crustal thickness of 49 ± 16 km. The crustal structure has also been locally constrained by seismic means, primarily beneath the Apollo 12 and 14 stations (see also Chenet *et al.*, 2006), but these investigations are not entirely in agreement. Initially, a value of about 60 km was reported by Toksöz *et al.* (1974), but more recent studies imply thinner values of 45 ± 5 km (Khan *et al.*, 2000), 38 ± 8 (Khan and Mosegaard, 2002), and 30 ± 2.5 km (Lognonné *et al.*, 2003). When the elevations of the Apollo stations are taken into account, the measured GTRs are most consistent with the recent thin-crust seismic estimates.

By assuming values for the mean crustal thickness, as well as the density of the crust and mantle, it is possible to construct a global crustal thickness model of the Moon (see Wieczorek and Phillips, 1998; Wieczorek *et al.*, 2006, and Section 10.05.5). The most notable feature of these models is the dramatic thinning of the crust beneath the large impact basins. This is a natural consequence of the large quantity of material that is ballistically excavated during the impact process (e.g., Wieczorek and Phillips, 1999), and it is seen that the depth of excavation reaches several tens of kilometers beneath the largest basins. It seems probable that some basins, such as Crisium, might even have excavated into the underlying mantle given their inferred near-zero crustal thicknesses. Nevertheless, despite the great size of the giant South Pole-Aitken basin on the farside, it appears that its depth of excavation was relatively shallow, and that \sim 20 km of crustal materials are present there. If the assumption of a constant density crust is correct, then the \sim 1.9 km center-of-mass/center-of-figure offset implies that the farside crust is thicker than the nearside hemisphere by about 15 km. However, global scale variations in crustal composition are known to

exist (see Jolliff *et al.*, 2000), and if these affect the crustal density, the hemispheric difference in crustal thickness could be much less.

The largest lunar impact basins (excluding the South Pole-Aitken basin) are characterized by having low elevations and large positive gravity anomalies, a signature generally referred to as a ‘mascon’ basin. The positive gravitational anomalies are likely a result of both uplift of the underlying crust–mantle interface, and the flexural support of surface mare basalt flows. Based on estimates of the mare basalt thicknesses, which can reach a few kilometers within the central portions of some impact basins, it appears that the crust–mantle interface has, in some cases, been uplifted above its pre-mare isostatic position (e.g., Neumann *et al.*, 1996; Wieczorek and Phillips, 1999). This hypothesis is supported by the existence of mascon basins that lack evidence of mare volcanism (Konopliv *et al.*, 1998). Those basins that are in a pre-mare isostatic state appear to be confined to a region of the crust that is enhanced in heat producing elements, and which likely possesses higher temperatures (Wieczorek and Phillips, 1999, 2000). In contrast to the largest impact basins, intermediate-sized craters have negative gravity anomalies and generally show some form of compensation (e.g., Reindler and Arkani-Hamed, 2001). Only about 15% of the craters in the Reindler and Arkani-Hamed (2001) study appear to be completely uncompensated, or to possess excess negative gravity anomalies due to crustal brecciation. For these intermediate sized craters, there does not appear to be any correlation of compensation state with crater age or location.

A few studies have attempted to place constraints on the elastic thickness of the Moon using both gravity and topography data (e.g., Arkani-Hamed, 1998; Crosby and McKenzie, 2005; Sugano and Heki, 2004). Most analyses have concentrated on the mascon basins, but unfortunately, the validity of the employed assumptions is often difficult to quantify. A proper analysis requires an assessment of (1) whether or not the mascon basins were in an isostatic (or super-isostatic) state before they were loaded by mare basalts, (2) the geometry and thickness of the mare basalt loads, (3) finite amplitude contributions of the uplifted crust–mantle interface, (4) both the admittance and correlation functions if the analysis is performed in the spectral domain, (5) a loading model that takes account of the surface and subsurface loads and their unknown phase relationship, and (6) the proper diameter of a basin (‘main topographic

rims’ often differ significantly from the more relevant diameter of the excavation cavity (Wieczorek and Phillips, 1999)). An alternative method for estimating the elastic thickness is by comparing the location of tectonic features (such as faults and graben) to that predicted by a specified loading model (e.g., Solomon and Head, 1980). Detailed modeling of the Serenitatis basin (Freed *et al.*, 2001) suggests that its elastic thickness was about 25 km when the concentric rilles formed, and probably greater than 70 km when the younger compressional ridges formed.

Finally, one curious large-scale feature of Moon is the amplitude of its degree-2 gravity and topography terms. If the Moon were in hydrostatic equilibrium, the amplitude of the C_{20} and C_{22} terms would be directly relatable to the Earth-Moon separation (cf. eqn [18]). The present day magnitudes of these coefficients, however, are much greater than would be expected for equilibrium at the present time. This has led to the suggestion that the equilibrium shape of the Moon was frozen into the lithosphere when it was closer to the Earth early in its orbital evolution (e.g., Jeffreys, 1976). If the observed magnitudes are interpreted as a relict equilibrium shape, then the corresponding Earth-Moon separation is about 25 Earth radii (the current separation is about 60 Earth radii) (Lambeck and Pullan, 1980). This interpretation is somewhat problematical as the lunar orbit is predicted to have receded beyond this distance in less than \sim 100 My after the formation of the Earth-Moon system (e.g., Webb, 1982). Alternatively, it is possible that this shape is a result of large-scale crustal thickness variations, or lateral variations in mantle density.

10.05.9 Future Developments and Concluding Remarks

The gravitational fields and topography of the terrestrial planets have become increasingly better characterized since the discovery of lunar ‘mascons’ by Muller and Sjogren in 1968. While the early data sets were quite sparse, the gradual accumulation of data with each successive space mission have given rise to near-global gravity and topography spherical harmonic models. Some of the gravity models now possess spherical harmonic bandwidths greater than 100, and future missions will surely lead to vast improvements. With the exception of the Moon, the topography has been measured to an accuracy that exceeds that of the corresponding gravity model.

Not only has the resolution of the planetary data sets continued to improve with time, but so have the analysis techniques. Early investigations were often restricted to analyses of individual 1-D line-of-sight gravitational acceleration profiles. As data coverage became more dense, 2-D regional models were developed that were more often than not analyzed using Cartesian techniques developed for the Earth. Because of the small size of some planetary bodies, such as the Moon, the assumption of Cartesian geometry has been called into question, and spherical analysis methods have proven to be superior. In the past ten years, the full suite of Cartesian gravity-topography analysis techniques have been developed for the sphere, including multitaper spectral analysis, the rapid calculation of gravitational anomalies from finite amplitude topographic relief, and realistic admittance models that take into account surface and subsurface loading with an arbitrary phase relationship. Though the approximation of Cartesian geometry may not incur large errors for some small-scale investigations, it is currently just as easy to use a spherical-based method that possess a comparable computational speed.

While much has been learned about the crustal and lithospheric structure of the terrestrial planets, there is still much to be done. In particular, in hindsight it is now clear that many gravity-topography admittance and coherence studies have used analysis techniques that can yield unreliable results. Shortcomings include incorrect application of the multitaper spectral analysis technique, the neglect of either the admittance or correlation function, and the use of a theoretical admittance model that might be an oversimplification of reality. Few studies, even for the Earth, have performed these analyses entirely correctly, and one should be quite skeptical of the majority of elastic thickness estimates that have been published for regions where subsurface loading is important.

It is also important to note that the concept of an ‘elastic’ lithosphere is in actuality a gross oversimplification of reality. Because of the ease of generating a time-invariant flexural profile from a load emplaced on an elastic plate, we would like to hope that the obtained ‘effective elastic thickness’ has some meaning. While this might be true for regions where the magnitude of surface and subsurface loading is small, and where inplane forces are absent, it has been demonstrated that the use of a more realistic rheology can yield flexural profiles that sometimes are quite different. Unfortunately, the most realistic EVP models are computationally expensive, and are not currently amenable to a robust inversion procedure

using gravity and topography as constraints. Though a simpler elastic-plastic formulation could be used in such an inversion, this rheological model utilizes assumptions that still might be too simplistic. Nevertheless, it would be appropriate to develop an elastic-plastic loading model similar to the elastic model described in Section 10.05.7. One benefit of such a model is that it would be possible to invert for the regional heat flow. An additional avenue of future research is to compare the locations of surface faulting with those predicted from elastic, elastic-plastic, and EVP models.

Finally, it is worth mentioning that significant improvements will be made to our knowledge of the gravitational fields and topography of the terrestrial planets. In particular, while the land-based topography for the Earth is now known to high accuracy, there are still gaps near the polar regions that could be partially filled by data obtained by the orbiting GLAS laser altimeter (e.g., Abshire *et al.*, 2005; Schutz *et al.*, 2005; Shuman *et al.*, 2006). The soon to be launched satellite GOCE, which contains a gravity gradiometer, will lead to improved models of the terrestrial gravity field. An area of active research for the Earth, but also for the other planets in a more limited sense, is that of measuring and modeling time variable gravity signatures that are a result of hydrologic processes and tidal deformation.

In addition to the Earth, spacecraft missions to the other terrestrial planets are bound to yield surprises. The lunar topography will be dramatically improved by the upcoming SELENE and Lunar Reconnaissance Orbiter (LRO) missions. Analysis of data obtained from the SELENE relay satellite, as well as dense altimetric cross-overs from LRO, will furthermore vastly improve our knowledge of the Moon’s farside gravity field. The gravity field of Mars will be improved by data obtained from the currently orbiting Mars Reconnaissance Orbiter. The topography and gravity field of Mercury will be characterized globally for the first time from the MESSENGER (Solomon *et al.*, 2001) and BepiColumbo (Spohn *et al.*, 2001) missions. In addition, missions are currently being proposed to measure the gravity and topography of bodies in the outer solar system, such as the satellites of Jupiter and Saturn.

Acknowledgments

Author is indebted to Greg Neumann and Frederik Simons for numerous discussions relating to the more

subtle details of gravity and topography acquisition and analysis. Francis Nimmo, Bruce Bills, David Sandwell and Martin Power are kindly thanked for constructive reviews and discussions that helped improve this manuscript.

All figures were created using the Generic Mapping Tools of Wessel and Smith (1991).

References

- Abe S, Mukai T, Hirata N, et al. (2006) Mass and local topography measurements of Itokawa by Hayabusa. *Science* 312(5776): 1344–1347 (doi:10.1126/science.1126272).
- Abshire JB, Sun X, Riris H, et al. (2005) Geoscience laser altimeter system (GLAS) on the ICESat mission: On-orbit measurement performance. *Geophysical Research Letters* 32 (L21S02), doi:10.1029/2005GL024028.
- Albert R and Phillips RJ (2000) Paleoflexure. *Geophysical Research Letters* 27: 2385–2388.
- Albert R, Phillips RJ, Dombard A, and Brown C (2000) A test of the validity of yield strength envelopes with an elastovisco-plastic finite element model. *Geophysical Journal International* 140: 399–409.
- Anderson JD, Colombo G, Esposito PB, Lau EL, and Trager GB (1987) The mass, gravity field, and ephemeris of Mercury. *Icarus* 71: 337–349.
- Anderson JD, Jacobson RA, Lau EL, Moore WB, and Schubert G (2001a) Io's gravity field and interior structure. *Journal of Geophysical Research* 106: 32963–32969.
- Anderson JD, Jacobson RA, McElrath TP, Moore WB, and Schubert G (2001b) Shape, mean radius, gravity field, and interior structure of Callisto. *Icarus* 153: 157–161.
- Anderson JD, Jurgens RF, Lau EL, Slade MA, III, and Schubert G (1996a) Shape and orientation of Mercury from radar ranging data. *Icarus* 124: 690–697.
- Anderson JD, Lau EL, Sjogren WL, Schubert G, and Moore WB (1996b) Gravitational constraints on the internal structure of Ganymede. *Nature* 384: 541–543.
- Anderson JD, Schubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998) Europa's differentiated internal structure: Inferences from four Galileo encounters. *Science* 281: 2019–2022.
- Anderson JD, Sjogren WL, and Schubert G (1996c) Galileo gravity results and the internal structure of Io. *Science* 272: 709–712.
- André SL, Watters TR, and Robinson MS (2005) The long wavelength topography of Beethoven and Tolstoj basins, Mercury. *Geophysical Research Letters* 32 (L21202), doi:10.1029/2005GL023627.
- Arkani-Hamed J (1998) The lunar mascons revisited. *Journal of Geophysical Research* 103: 3709–3739.
- Arkani-Hamed J (2000) Strength of Martian lithosphere beneath large volcanoes. *Journal of Geophysical Research* 105: 26713–26732.
- Balmino G (1993) The spectra of the topography of the Earth, Venus and Mars. *Geophysical Research Letters* 20(11): 1063–1066.
- Balmino G (1994) Gravitational potential harmonics from the shape of an homogeneous body. *Celestial Mechanics and Dynamical Astronomy* 60: 331–364.
- Banerdt WB (1986) Support of long-wavelength loads on Venus and implications for internal structure. *Journal of Geophysical Research* 91: 403–419.
- Banerdt WB, Golombek M, and Tanaka K (1992) Stress and tectonics on Mars. In: Kieffer H, Jakosky B, Snyder C, and Matthews M (eds.) *Mars*, pp. 249–297. Tucson: University of Arizona Press.
- Banks RJ, Francis SC, and Hipkin RG (2001) Effects of loads in the upper crust on estimates of the elastic thickness of the lithosphere. *Geophysical Journal International* 145: 291–299.
- Barnett DN, Nimmo F, and McKenzie D (2000) Elastic thickness estimates for Venus using line of sight accelerations from Magellan cycle 5. *Icarus* 146: 404–419.
- Barnett DN, Nimmo F, and McKenzie D (2002) Flexure of Venusian lithosphere measured from residual topography and gravity. *Journal of Geophysical Research* 107 (E2), doi:10.1029/2000JE001398.
- Bechtel TD, Forsyth DW, Sharpton VL, and Grieve RAF (1990) Variations in effective elastic thickness of the North-American lithosphere. *Nature* 343(6259): 636–638.
- Bechtel TD, Forsyth DW, and Swain CJ (1987) Mechanisms of isostatic compensation in the vicinity of the East African Rift, Kenya. *Geophysical Journal of the Royal Astronomical Society* 90: 445–465.
- Belleguic V, Lognonné P, and Wieczorek M (2005) Constraints on the Martian lithosphere from gravity and topography data. *Journal of Geophysical Research* 110 (E11005), doi:10.1029/2005JE002437.
- Bills BG and Lemoine FG (1995) Gravitational and topographic isotropy of the Earth, Moon, Mars, and Venus. *Journal of Geophysical Research* 100: 26275–26295.
- Blakely RJ (1995) *Potential Theory in Gravity and Magnetic Applications*. New York: Cambridge University Press.
- Brown CD and Grimm RE (1996) Lithospheric rheology and flexure at Artemis Chasma, Venus. *Journal of Geophysical Research* 101(E5): 12697–12708.
- Brown CD and Phillips RJ (2000) Crust–mantle decoupling by flexure of continental lithosphere. *Journal of Geophysical Research* 105(B6): 13221–13237.
- Burov EB and Diamant M (1995) The effective elastic thickness (T_e) of continental lithosphere: What does it really mean? *Journal of Geophysical Research* 100(B3): 3905–3927.
- Ceuleneer G, Rabinowicz M, Monnereau M, Cazenave A, and Rosemburg C (1988) Viscosity and thickness of the sub-lithospheric low-viscosity zone: Constraints from geoid and depth over oceanic swells. *Earth and Planetary Science Letters* 89: 84–102.
- Chambat F and Valette B (2005) Earth gravity up to second order in topography and density. *Physics of the Earth and Planetary Interiors* 151: 89–106.
- Chao BF and Rubincam DP (1989) The gravitational field of Phobos. *Geophysical Research Letters* 16: 859–862.
- Chenet H, Lognonné P, Wieczorek M, and Mizutani H (2006) Lateral variations of lunar crustal thickness from the Apollo seismic data set. *Earth and Planetary Science Letters* 243: 1–14.
- Cheng AF, Barnouin-Jha O, Prockter L, et al. (2002) Small-scale topography of 433 Eros from laser altimetry and imaging. *Icarus* 155: 51–74.
- Cook AC, Watters TR, Robinson MS, Spudis PD, and Bussey DBJ (2000) Lunar polar topography derived from Clementine stereoimages. *Journal of Geophysical Research* 105: 12023–12033.
- Crosby A and McKenzie D (2005) Measurements of the elastic thickness under ancient lunar terrain. *Icarus* 173: 100–107.
- Dahlen FA and Tromp J (1998) *Theoretical Global Seismology*. Princeton, NJ: Princeton University Press.
- Dermott SF (1979) Shapes and gravitational moments of satellites and asteroids. *Icarus* 37: 575–586.
- Dorman LM and Lewis BT (1970) Experimental isostasy. 1: Theory of the determination of the Earth's isostatic response

- to a concentrated load. *Journal of Geophysical Research* 75(17): 3357–3365.
- Driscoll JR and Healy DM (1994) Computing Fourier transforms and convolutions on the 2-sphere. *Advances in Applied Mathematics* 15: 202–250.
- Duxbury TC (1989) The figure of Phobos. *Icarus* 78: 169–180.
- Ebinger C, Bechtel T, Forsyth D, and Bowin C (1989) Effective elastic plate thickness beneath the east African and Afar plateaus and dynamic compensation of the uplifts. *Journal of Geophysical Research* 94(B3): 2883–2901.
- Esposito P, Benerdt W, Lindal G, Sjogren W, and Slade M (1992) Gravity and topography. In: Kieffer H, Jakosky B, Snyder C, and Matthews M (eds.) *Mars*, pp. 209–248. Tucson: University of Arizona Press.
- Fedi M and Florio G (2002) A stable downward continuation by using the ISVD method. *Geophysical Journal International* 151: 146–156.
- Ford P and Pettengill G (1992) Venus topography and kilometer-scale slopes. *Journal of Geophysical Research* 97: 13103–13114.
- Forsyth DW (1985) Subsurface loading and estimates of the flexural rigidity of continental lithosphere. *Journal of Geophysical Research* 90(B14): 12623–12632.
- Freed AM, Melosh HJ, and Solomon SC (2001) Tectonics of mascon loading: Resolution of the strike-slip faulting paradox. *Journal of Geophysical Research* 106: 20603–20620.
- Garmier R and Barriot J-P (2001) Ellipsoidal harmonic expansions of the gravitational potential: Theory and application. *Celestial Mechanics and Dynamical Astronomy* 79: 235–275.
- Garmier R and Barriot J-P (2002) Modeling of the Eros gravity field as an ellipsoidal harmonic expansion from the NEAR doppler tracking data. *Geophysical Research Letters* 29: 72-1–72-3 (doi:10.1029/2001GL013768).
- Ghent ED, Dipple GM, and Russell JK (2004) Thermodynamic models for eclogite mantle lithosphere. *Journal of Geophysical Research* 218: 451–462.
- Giese B, Oberst J, Roatsch T, Neukum G, Head JW, and Pappalardo R (1998) The local topography of Uruk Sulcus and Galileo Regio obtained from stereo images. *Icarus* 135: 303–316.
- Goossens S, Visser PNAM, and Ambrosius BAC (2005a) A method to determine regional lunar gravity fields from earth-based satellite tracking data. *Planetary and Space Science* 53: 1331–1340.
- Goossens S, Visser PNAM, Heki K, and Ambrosius BAC (2005b) Local gravity from Lunar Prospector tracking data: Results of Mare Serenitatis. *Earth, Planets and Space* 57: 1127–1132.
- Grimm RE (1994) The deep structure of Venusian plateau highlands. *Icarus* 112: 89–103.
- Grimm RE and Hess PC (1997) The crust of Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1205–1244. Tucson: University of Arizona Press.
- Grimm RE and Phillips RJ (1991) Gravity anomalies, compensation mechanisms, and the geodynamics of Western Ishtar Terra, Venus. *Journal of Geophysical Research* 96(B6): 8309–8324.
- Haines GV (1985) Spherical cap harmonic analysis. *Journal of Geophysical Research* 90: 2583–2591.
- Hansen VL and Phillips RJ (1995) Formation of Ishtar Terra, Venus: Surface and gravity constraints. *Geology* 23: 292–296.
- Harmon JK, Campbell DB, Bindschadler DL, Head JW, and Shapiro II (1986) Radar altimetry of Mercury: A preliminary analysis. *Journal of Geophysical Research* 91: 385–401.
- Hastings DA and Dunbar PK (1999) *Global Land One-kilometer Base Elevation (GLOBE) Digital Elevation Model, Documentation, Volume 1.0. Key to Geophysical Records Documentation (KGRD)* 34. Boulder, CO, USA: National Oceanic and Atmospheric Administration, National Geophysical Data Center.
- Hauck SA, III, Phillips RJ, and Price MH (1998) Venus: Crater distribution and plains resurfacing models. *Journal of Geophysical Research* 103(E6): 13635–13642.
- Haxby WF and Turcotte DL (1978) On isostatic geoid anomalies. *Journal of Geophysical Research* 83: 5473–5478.
- Heiskanen WA and Moritz H (1967) *Physical Geodesy*. San Francisco: Freeman.
- Holmes SA and Featherstone WE (2002) A unified approach to the Clenshaw summation and the recursive computation of very high degree and order normalised associated Legendre functions. *Journal of Geodesy* 76: 279–299.
- Hoogenboom T, Smrekar SE, Anderson FS, and Houseman G (2004) Admittance survey of type 1 coronae on Venus. *Journal of Geophysical Research* 109 (E03002), doi:10.1029/2003JE002171.
- Hoogenboom T, Houseman TG, and Martin P (2005) Elastic thickness estimates for coronae associated with chasmata on Venus. *Journal of Geophysical Research* 110 (E09003), doi:10.1029/2004JE002394.
- Hwang C and Chen S-K (1997) Fully normalized spherical cap harmonics: Application to the analysis of sea-level data from TOPEX/POSEIDON and ERS-1. *Geophysical Journal International* 129: 450–460.
- Hwang C and Parsons B (1996) An optimal procedure for determining gravity from multi-satellite altimetry. *Geophysical Journal International* 125: 705–719.
- Jeffreys H (1976) *The Earth: Its Origin History and Physical Constitution* 6th edn. Cambridge: Cambridge University Press.
- Johnson C, Solomon S, Head JW, III, Phillips RJ, Smith D, and Zuber M (2000) Lithospheric loading by the northern polar cap on Mars. *Icarus* 144: 313–328.
- Johnson CL and Sandwell DT (1994) Lithospheric flexure on Venus. *Geophysical Journal International* 119: 627–647.
- Jolliff BJ, Gillis JJ, Haskin LA, Korotev RL, and Wieczorek MA (2000) Major lunar crustal terranes: Surface expressions and crust-mantle origins. *Journal of Geophysical Research* 105: 4197–4216.
- Kaula WM (1967) Theory of statistical analysis of data distributed over a sphere. *Reviews of Geophysics* 5(1): 83–107.
- Kaula WM (2000) *Theory of Satellite Geodesy: Applications of Satellites to Geodesy*. Mineola, NY: Dover.
- Khan A and Mosegaard K (2002) An enquiry into the lunar interior – A non-linear inversion of the Apollo lunar seismic data. *Journal of Geophysical Research* 107(E6): 3-1–3-18 (10.1029/2001JE001,658).
- Khan A, Mosegaard K, and Rasmussen KKL (2000) A new seismic velocity model for the Moon from a Monte Carlo inversion of the Apollo lunar seismic data. *Geophysical Research Letters* 27: 1591–1594.
- Kiefer W (2004) Gravity evidence for an extinct magma chamber beneath Syrtis Major, Mars: A look at the magmatic plumbing system. *Earth and Planetary Science Letters* 222: 349–361.
- Kiefer W (2005) Buried mass anomalies along the hemispheric dichotomy in eastern Mars: Implications for the origin and evolution of the dichotomy. *Geophysical Research Letters* 32 (L22201), doi:10.1029/2005GL024260.
- Kiefer WS and Häger BF (1991) A mantle plume model for the equatorial highlands of Venus. *Journal of Geophysical Research* 96(E4): 20947–20966.
- Kiefer WS, Richards MA, and Häger BH (1986) A dynamic model of Venus's gravity field. *Geophysical Research Letters* 13(1): 14–17.
- Konopliv AS, Asmar SW, and Yuan DN (2001) Recent gravity models as a result of the Lunar Prospector mission. *Icarus* 150: 1–18.
- Konopliv AS, Banerdt WB, and Sjogren WL (1999) Venus gravity: 180th degree and order model. *Icarus* 139: 3–18.

- Konopliv AS, Binder AB, Hood LL, Kucinskas AB, Sjogren WL, and Williams JG (1998) Improved gravity field of the Moon from Lunar Prospector. *Science* 281: 1476–1480.
- Konopliv AS, Yoder CF, Standish EM, Yuan D-N, and Sjogren WL (2006) A global solution for the Mars static and seasonal gravity, Mars orientation, Phobos and Deimos masses, and Mars ephemeris. *Icarus* 182: 23–50.
- Kraus H (1967) *Thin Elastic Shells: An Introduction to the Theoretical Foundations and the Analysis of Their Static and Dynamic Behavior*, 467pp. New York: John Wiley.
- Kucinskas AB and Turcotte DT (1994) Isostatic compensation of equatorial highlands on Venus. *Icarus* 112: 104–116.
- Lambeck K (1988) *Geophysical Geodesy*. New York: Oxford University Press.
- Lambeck K and Pullan S (1980) The lunar fossil bulge hypothesis revisited. *Physics of the Earth and Planetary Interiors* 22: 29–35.
- Lawrence KP and Phillips RJ (2003) Gravity/topography admittance inversion on Venus using niching genetic algorithms. *Geophysical Research Letters* 30(19): 1994 (doi:10.1029/2003GL017515).
- Lemoine FG, Kenyon SC, Foutor JK, et al. (1998) *The Development of the Joint NASA GSFC and the National Imagery and Mapping Agency (NIMA) Geopotential Model EGM96*. Greenbelt, MD: NASA Goddard Space Flight Center.
- Lemoine FG, Smith DE, Rowlands DD, et al. (2001) An improved solution of the gravity field of Mars (GMM-2B) from Mars Global Surveyor. *Journal of Geophysical Research* 106(E10): 23359–23376.
- Lognonné P, Gagnepain-Beyneix J, and Chenet H (2003) A new seismic model for the Moon: Implications for structure, thermal evolution and formation of the Moon. *Earth and Planetary Science Letters* 211: 27–44.
- Lowry AR and Zhong S (2003) Surface versus internal loading of the Tharsis rise, Mars. *Journal of Geophysical Research* 5099(E9), doi:10.1029/2003JE002111.
- Margot J-L, Campbell DB, Jurgens RF, and Slade MA (1999a) Topography of Tycho crater. *Journal of Geophysical Research* 104: 11875–11882.
- Margot J-L, Campbell DB, Jurgens RF, and Slade MA (1999b) Topography of the lunar poles from radar interferometry: A survey of cold trap locations. *Science* 284: 1658–1660.
- Marks KM and Smith WHF (2006) An evaluation of publicly available global bathymetry grids. *Marine Geophysical Research* 27: 19–34.
- Martinec Z, Pěč K, and Burša M (1989) The Phobos gravitational field modeled on the basis of its topography. *Earth, Moon and Planets* 45: 219–235.
- McGovern PJ, Solomon SC, Smith DE, et al. (2002) Localized gravity/topography admittance and correlation spectra on Mars: Implications for regional and global evolution. *Journal of Geophysical Research* 107(E12): 5136 (doi:10.1029/2002JE001854).
- McGovern PJ, Solomon SC, Smith DE, et al. (2004) Correction to ‘Localized gravity/topography admittance and correlation spectra on Mars: Implications for regional and global evolution’. *Journal of Geophysical Research* 109(E07007), doi:10.1029/2004JE002286.
- McKenzie D (1994) The relationship between topography and gravity on Earth and Venus. *Icarus* 112: 55–88.
- McKenzie D (2003) Estimating T_e in the presence of internal loads. *Journal of Geophysical Research* 108(B9): 2348 (doi:10.1029/2002JB001766).
- McKenzie D and Nimmo F (1997) Elastic thickness estimates for Venus from line of sight accelerations. *Icarus* 130: 198–216.
- McKenzie DP, Barnett DN, and Yuan D-N (2002) The relationship between Martian gravity and topography. *Earth and Planetary Science Letters* 195: 1–16.
- McNutt MK (1984) Lithospheric flexure and thermal anomalies. *Journal of Geophysical Research* 89(B13): 11180–11194.
- Miller JK, Konopliv AS, Antreasian, et al. (2002) Determination of shape, gravity, and rotational state of asteroid 433 Eros. *Icarus* 155: 3–17.
- Mohr PJ and Taylor BN (2005) CODATA recommended values of the fundamental physical constants: 2002. *Reviews of Modern Physics* 77: 1–107.
- Moore WB and Schubert G (1995) Lithospheric thickness and mantle/lithosphere density contrast beneath Beta Regio, Venus. *Geophysical Research Letters* 22(4): 429–432.
- Moore WB and Schubert G (1997) Venusian crustal and lithospheric properties from nonlinear regressions of highland geoid and topography. *Icarus* 128: 415–428.
- Moresi L and Solomatov V (1998) Mantle convection with a brittle lithosphere: Thoughts on the global tectonic styles of the Earth and Venus. *Geophysical Journal International* 133: 669–682.
- Muller PM and Sjogren WL (1968) Masons: Lunar mass concentrations. *Science* 161: 680–684.
- Mueller S and Phillips RJ (1995) On the reliability of lithospheric constraints derived from models of outer-rise flexure. *Geophysical Journal International* 123: 887–902.
- National Geophysical Data Center (2001) *ETOPO2 Global 2' Elevations [CD-ROM]*. Boulder, CO, USA: US Department of Commerce, National Oceanic and Atmospheric Administration.
- National Imagery and Mapping Agency (2000) *Department of Defense World Geodetic System 1984: Its Definition and Relationship with Local Geodetic Systems*, NIMA TR8350.2, 175pp. Bethesda, MD: National Imagery and Mapping Agency.
- Neumann GA, Rowlands DD, Lemoine FG, Smith DE, and Zuber MT (2001) Crossover analysis of Mars Orbiter Laser Altimeter data. *Journal of Geophysical Research* 106: 23753–23768.
- Neumann GA, Zuber MT, Smith DE, and Lemoine FG (1996) The lunar crust: Global structure and signature of major basins. *Journal of Geophysical Research* 101: 16841–16843.
- Neumann GA, Zuber MT, Wieczorek MA, McGovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research* 109: E08002, doi:10.1029/2004JE002262.
- Nimmo F (2002) Admittance estimates of mean crustal thickness and density at the Martian hemispheric dichotomy. *Journal of Geophysical Research* 107(E11): 5117 (doi:10.1029/2000JE001488).
- Nimmo F, Giese B, and Pappalardo RT (2003a) Estimates of Europa’s ice shell thickness from elastically-supported topography. *Geophysical Research Letters* 30: 1233 (doi:10.1029/2002GL016660).
- Nimmo F and McKenzie D (1998) Volcanism and tectonics on Venus. *Annual Reviews of Earth and Planetary Science* 26: 23–51.
- Nimmo F, Pappalardo RT, and Giese G (2003b) On the origins of band topography, Europa. *Icarus* 166: 21–32.
- Nimmo F and Stevenson DJ (2001) Estimates of Martian crustal thickness from viscous relaxation of topography. *Journal of Geophysical Research* 106: 5085–5098.
- Nimmo F and Tanaka K (2005) Earth crustal evolution of Mars. *Annual Reviews of Earth and Planetary Science* 33: 133–161 (doi:10.1146/annurev.earth.33.092203.122,637).
- Ockendon JR and Turcotte DL (1977) On the gravitational potential and field anomalies due to thin mass layers. *Geophysical Journal of the Royal Astronomical Society* 48: 479–492.
- Panasuk SV and Hager BH (2000) Inversion for mantle viscosity profiles constrained by dynamic topography and the geoid, and their estimated errors. *Geophysical Journal International* 143: 821–836.

- Parker RL (1972) The rapid calculation of potential anomalies. *Geophysical Journal of the Royal Astronomical Society* 31: 447–455.
- Parmetier EM and Hess PC (1992) Chemical differentiation of a convecting planetary interior: Consequences for a one plate planet such as Venus. *Geophysical Research Letters* 19(20): 2015–2018.
- Pauer M, Fleming K, and Čadek O (2006) Modeling the dynamic component of the geoid and topography of Venus. *Journal of Geophysical Research* 111: E11012 (doi:10.1029/2005JE002511).
- Percival DB and Walden AT (1993) *Spectral Analysis for Physical Applications, Multitaper and Conventional Univariate Techniques*. New York: Cambridge University Press.
- Pérez-Gussinyé M, Lowry A, Watts AB, and Velicogna I (2004) On the recovery of effective elastic thickness using spectral methods: Examples from synthetic data and from the Fennoscandian Shield. *Journal of Geophysical Research* 109 (B10409), doi:10.1029/2003JB002788.
- Phillips RJ (1994) Estimating lithospheric properties at Atla Regio. *Icarus* 112: 147–170.
- Phillips RJ, Johnson CL, Mackwell SJ, Morgan P, Sandwell DT, and Zuber MT (1997) Lithospheric mechanics and dynamics of Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1163–1204. Tucson: University of Arizona Press.
- Phillips RJ and Lambeck K (1980) Gravity fields of the terrestrial planets: Long-wavelength anomalies and tectonics. *Reviews of Geophysics* 18(1): 27–76.
- Phillips RJ, Zuber MT, Solomon SC, et al. (2001) Ancient geo-dynamics and global-scale hydrology on Mars. *Science* 291: 2587–2591.
- Phipps Morgan J and Blackman DK (1993) Inversion of combined gravity and bathymetry data for crustal structure: A prescription for downward continuation. *Earth and Planetary Science Letters* 119: 167–179.
- Press WH, Teukolsky SA, Vetterling WT, and Flannery BP (1992) *Numerical Recipes in FORTRAN: The Art of Scientific Computing* 2nd edn. Cambridge: Cambridge University Press.
- Rabus PB, Eineder M, Roth A, and Bamler R (2003) The shuttle radar topography mission – A new class of digital elevation models acquired by spaceborne radar. *Photogrammetric Remote Sensory* 57: 241–262.
- Rapp RH (1989) The decay of the spectrum of the gravitational potential and the topography of the Earth. *Geophysical Journal International* 99: 449–455.
- Rappaport NJ, Konopliv AS, and Kucinskas AB (1999) An improved 360 degree and order model of Venus topography. *Icarus* 139: 19–31.
- Reigber C, Jochmann H, Wünsch, et al. (2004) Earth gravity field and seasonal variation from CHAMP. In: Reigber C, Lühr H, Schwintzer P, and Wickert J (eds.) *Earth Observations with CHAMP*, pp. 25–30. Berlin: Springer.
- Reigber C, Schmidt R, Flechtner F, et al. (2005) An Earth gravity field model complete to degree and order 150 from GRACE: EIGEN-GRACE02S. *Journal of Geodynamics* 39: 1–10.
- Reinhard L and Arkani-Hamed J (2001) The compensation state of intermediate size lunar craters. *Icarus* 153: 71–88.
- Roberts J and Zhong S (2004) Plume-induced topography and geoid anomalies and their implications for the Tharsis rise on Mars. *Journal of Geophysical Research* 109 (E03009), doi:10.1029/2003JE002226.
- Robinson EM and Parsons B (1988) Effect of a shallow low-viscosity zone on the formation of midplate swells. *Journal of Geophysical Research* 93(B4): 3144–3156.
- Rodríguez E, Morris CS, Belz JE, et al. (2005) An assessment of the SRTM topographic products. *Technical Report of the Jet Propulsion Laboratory*. Pasadena, CA: NASA and CalTech.
- Rogers PG and Zuber MT (1998) Tectonic evolution of Bell Regio, Venus: Regional stress, lithospheric flexure, and edifice stresses. *Journal of Geophysical Research* 103(E7): 16841–16853.
- Rosenblatt P, Pinet P, and Touvenot E (1994) Comparative hypsometric analysis of Earth and Venus. *Geophysical Research Letters* 21: 465–468.
- Rosiek MR and Aeschliman R (2001) Lunar shaded relief map updated with Clementine data. *Lunar and Planetary Science XXXII (CD-ROM)*, 1943.
- Rummel R (2005) Gravity and topography of Moon and planets. *Earth, Moon and Planets* 94: 103–111 (doi:10.1007/s11038-005-3245-z).
- Sandwell DT and Schubert G (1992) Flexural ridges, trenches, and outer rises around coronae on Venus. *Journal of Geophysical Research* 97(E10): 16069–16083.
- Sandwell DT and Smith WHF (1997) Marine gravity anomaly from Geosat and ERS 1 satellite altimetry. *Journal of Geophysical Research* 102(B5): 10039–10054.
- Scheeres DJ, Khushalani B, and Werner RA (2000) Estimating asteroid density distributions from shape and gravity information. *Planetary and Space Science* 48: 965–971.
- Schubert G and Sandwell DT (1995) A global survey of possible subduction sites on Venus. *Icarus* 117: 173–196.
- Schubert G, Solomatov VS, Tackley PJ, and Turcotte DL (1997) Mantle convection and the thermal evolution of Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1245–1287. Tucson: University of Arizona Press.
- Schutz BE, Zwally HJ, Shuman CA, Hancock D, and Di-Marzio JP (2005) Overview of the ICESat mission. *Geophysical Research Letters* 32 (L21S01), doi:10.1029/2005GL024009.
- Shuman C, Zwally HJ, Schutz BE, et al. (2006) ICESat Antarctic elevation data: Preliminary precision and accuracy assessment. *Geophysical Research Letters* 33 (L07501), doi:10.1029/2005GL025227.
- Simons FJ, Dahlen FA, and Wieczorek MA (2006) Spatiotemporal concentration on a sphere. *SIAM Review* 48(3): 504–536 (doi:10.1137/S003614450445765).
- Simons FJ, van der Hilst RD, and Zuber MT (2003) Spatiotemporal localization of isostatic coherence anisotropy in Australia and its relation to seismic anisotropy: Implications for lithospheric deformation. *Journal of Geophysical Research* 108(B5): 2250 (doi: 10.1029/2001JB000704).
- Simons FJ, Zuber MT, and Korenaga J (2000) Isostatic response of the Australian lithosphere: Estimation of effective elastic thickness and anisotropy using multitaper spectral analysis. *Journal of Geophysical Research* 105(B8): 19163–19184.
- Simons M, Hager BH, and Solomon SC (1994) Global variations in the geoid/topography admittance of Venus. *Science* 264: 798–803.
- Simons M, Solomon SC, and Hager BH (1997) Localization of gravity and topography: Constraints on the tectonics and mantle dynamics of Venus. *Geophysical Journal International* 131: 24–44.
- Sjogren WL, Banerdt WB, Chodas PW, et al. (1997) The Venus gravity field and other geodetic parameters. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1125–1161. Tucson: University of Arizona Press.
- Slepian D (1983) Some comments on Fourier analysis, uncertainty and modeling. *SIAM Reviews* 25(3): 379–393.
- Smith DE, Zuber MT, Frey HV, et al. (2001b) Mars Orbiter Laser Altimeter: Experiment summary after the first year of global mapping of Mars. *Geophysical Research Letters* 106(E10): 23689–23722.
- Smith DE, Zuber MT, Neumann GA, and Lemoine FG (1997) Topography of the Moon from Clementine lidar. *Journal of Geophysical Research* 102: 1591–1611.

- Smith DE, Zuber MT, and Neumann GA (2001a) Seasonal variations of snow depth on Mars. *Science* 294: 2141–2146.
- Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1503.
- Smith WHF and Sandwell DW (1994) Bathymetric prediction from dense satellite altimetry and sparse shipboard bathymetry. *Journal of Geophysical Research* 99: 21803–21824.
- Smith WHF and Sandwell DT (1997) Global sea floor topography from satellite altimetry and ship depth soundings. *Science* 277(5334): 1956–1962.
- Smrekar SE (1994) Evidence for active hotspots on Venus from analysis of Magellan gravity data. *Icarus* 112: 2–26.
- Smrekar SE, Comstock R, and Anderson FS (2003) A gravity survey of type 2 coronae on Venus. *Journal of Geophysical Research* 108(E8): 5090 (doi:10.1029/2002JE001935).
- Smrekar SE and Phillips RJ (1991) Venusian highlands: Geoid to topography ratios and their implications. *Earth and Planetary Science Letters* 107: 582–597.
- Smrekar SE and Stofan ER (1999) Origin of corona-dominated topographic rises on Venus. *Icarus* 139: 100–115.
- Sneeuw N (1994) Global spherical harmonic-analysis by least-squares and numerical quadrature methods in historical perspective. *Geophysical Journal International* 118(3): 707–716.
- Solomon SC and Head JW (1980) Lunar mascon basins: Lava filling, tectonics, and evolution of the lithosphere. *Reviews of Geophysics and Space Physics* 18: 107–141.
- Solomon SC and Head JW (1990) Lithospheric flexure beneath the Freya Montes foredeep; Venus: Constraints on lithospheric thermal gradient and heat flow. *Geophysical Research Letters* 17(9): 1393–1396.
- Solomon SC, McNutt RL, Jr., Gold RE, et al. (2001) MESSENGER mission to Mercury: Scientific objectives and implementation. *Planetary and Space Science* 49: 1445–1465.
- Spohn TF, Sohl K, Wieczorkowski V, and Conzelmann (2001) The interior structure of Mercury: What we know, what we expect from Bepicolobo. *Planetary and Space Science* 49: 1561–1570.
- Sugano T and Heki K (2004) Isostasy of the Moon from high-resolution gravity and topography data: Implication for its thermal history. *Geophysical Research Letters* 31 (L24703), doi:10.1029/2004GL022059.
- Tapley B, Ries J, Bettadpur S, et al. (2005) GGM02 – An improved Earth gravity field model from GRACE. *Journal of Geodesy* 79, doi:10.1007/s00190-005-0480-z.
- Tapley BD, Bettadpur S, Ries JC, Thompson PF, and Watkins M (2004) GRACE measurements of mass variability in the Earth system. *Science* 305: 503–505.
- Taylor GJ, Boynton W, Büchner J, et al. (2006) Bulk composition and early differentiation of Mars. *Journal of Geophysical Research* 111: E03510 (doi:10.1029/2005JE002645).
- Thébault E, Schott JJ, and Mandea M (2006) Revised spherical cap harmonic analysis (R-SCHA): Validation and properties. *Journal of Geophysical Research* 111 (B01102), doi:10.1029/2005JB003836.
- Thébault E, Schott JJ, Mandea M, and Hoffbeck JP (2004) A new proposal for spherical cap harmonic modelling. *Geophysical Journal International* 159: 83–103.
- Thomson DJ (1982) Spectrum estimation and harmonic analysis. *Proceedings of IEEE* 70(9): 1055–1096.
- Toksöz MN, Dainty AM, Solomon SC, and Anderson KR (1974) Structure of the moon. *Reviews of Geophysics* 12: 539–567.
- Turcotte DL (1995) How does Venus lose heat? *Journal of Geophysical Research* 100(E8): 16931–16940.
- Turcotte DL, Willemann RJ, Haxby WF, and Norberry J (1981) Role of membrane stresses in the support of planetary topography. *Journal of Geophysical Research* 86: 3951–3959.
- US Geological Survey (2002) Color-coded topography and shaded relief map of the lunar near side and far side hemispheres. I-2769.
- Varshalovich DA, Moskalev AN, and Khersonskii VK (1988) *Quantum Theory of Angular Momentum*. Singapore: World Scientific.
- Vezolainen AV, Solomatov VS, Basilevsky AT, and Head JW (2004) Uplift of Beta Regio: Three-dimensional models. *Journal of Geophysical Research* 109 (E08007), doi:10.1029/2004JE002259.
- Watters T (2003) Lithospheric flexure and the origin of the dichotomy boundary on Mars. *Geology* 31: 271–274.
- Watters TR, Cook AC, and Robinson MS (2001) Large-scale lobate scarps in the southern hemisphere of Mercury. *Planetary and Space Science* 49: 1523–1530.
- Watters TR, Schultz R, Robinson MS, and Cook AC (2002) The mechanical and thermal structure of Mercury's early lithosphere. *Geophysical Research Letters* 29(11): 37-1–37-4 (doi:10.1029/2001GL014308).
- Watts AB (2001) *Isostasy and Flexure of the Lithosphere*. Cambridge: Cambridge University Press.
- Watts AB and Burov EB (2003) Lithospheric strength and its relationship to the elastic and seismogenic layer thickness. *Earth and Planetary Science Letters* 213: 113–131.
- Watts AB and Zhong S (2000) Observations of flexure and the rheology of oceanic lithosphere. *Geophysical Journal International* 142: 855–875.
- Webb DJ (1982) Tides and the evolution of the Earth–Moon system. *Geophysical Journal of the Royal Astronomical Society* 70: 261–271.
- Werner RA (1997) Spherical harmonic coefficients for the potential of a constant-density polyhedron. *Journal of Comparative Geoscience* 23: 1071–1077.
- Werner RA and Scheeres DJ (1997) Exterior gravitation of a polyhedron derived and compared with harmonic and mascon gravitation representations of asteroid 4769 Castalia. *Celestial Mechanics and Dynamical Astronomy* 65: 313–344.
- Wessel P and Smith WHF (1991) Free software helps map and display data. *EOS Trans. AGU*, 72(41): 441 (doi:10.1029/90EO00319).
- Wieczorek M and Phillips RJ (1999) Lunar multiring basins and the cratering process. *Icarus* 139: 246–259.
- Wieczorek MA, Jolliff BL, Khan A, et al. (2006) The constitution and structure of the lunar interior. In: Jolliff BJ, Wieczorek MA, Shearer CK, and Neal CR (eds.) *New Views of the Moon, Reviews of Mineralogical Geochemistry*, vol. 60, pp. 221–364. Chantilly, VA: Mineralogical Society of America.
- Wieczorek MA (2007) The relationship between gravity and topography: Surface and subsurface loading of an elastic lithosphere. *Geophysics Journal International* (submitted).
- Wieczorek MA and Phillips RJ (1997) The structure and compensation of the lunar highland crust. *Journal of Geophysical Research* 102(E5): 10933–10943.
- Wieczorek MA and Phillips RJ (1998) Potential anomalies on a sphere: Applications to the thickness of the lunar crust. *Journal of Geophysical Research* 103(E1): 1715–1724.
- Wieczorek MA and Phillips RJ (2000) The ‘Procellarum KREEP Terrane’: Implications for mare volcanism and lunar evolution. *Journal of Geophysical Research* 105(E8): 20417–20430.
- Wieczorek MA and Zuber MT (2004) Thickness of the Martian crust: Improved constraints from geoid-to-topography ratios. *Journal of Geophysical Research* 109 (E01009), doi:10.1029/2003JE002153.
- Wieczorek MA and Simons FJ (2005) Localized spectral analysis on the sphere. *Geophysical Journal International* 162: 665–675 (doi:10.1111/j.1365–246X.2005.02687.X).
- Wieczorek MA and Simons FJ (2007) Minimum variance multi-taper spectral estimation on the sphere. *Journal of Fourier Analysis and Application*.

- Willemann RJ and Turcotte DL (1981) Support of topographic and other loads on the moon and on the terrestrial planets. *Proceedings of Lunar and Planetary Science Conference* 12B: 837–851.
- Yoder CF (1995) Astrometric and geodetic properties of Earth and the solar system. In: Ahrens TJ (ed.) *Global Earth Physics: A Handbook of Physical Constants*, AGU reference shelf 1, pp. 1–31. Washington, DC: American Geophysical Union.
- Yuan DN, Sjogren WL, Konopliv AS, and Kucinskas AB (2001) Gravity field of Mars: A 75th degree and order model. *Journal of Geophysical Research* 106(E10): 23377–23401.
- Zharkov VN, Leontjev VV, and Kozenko AV (1985) Models, figures and gravitational moments of the Galilean satellites of Jupiter and icy satellites of Saturn. *Icarus* 61: 92–100.
- Zhong S (2002) Effects of lithosphere on the long-wavelength gravity anomalies and their implications for the formation of the Tharsis rise on Mars. *Journal of Geophysical Research* 107(E7): 5054 (doi:10.1029/2001JE001589).
- Zuber MT (2001) The crust and mantle of Mars. *Nature* 412: 237–244.
- Zuber MT, Bechtel TD, and Forsyth DW (1989) Effective elastic thicknesses of the lithosphere and mechanisms of isostatic compensation in Australia. *Journal of Geophysical Research* 94(B7): 9353–9367.
- Zuber MT and Smith DE (1997) Mars without Tharsis. *Journal of Geophysical Research* 102: 28673–28685.
- Zuber MT, Smith DE, Andrew F, Cheng, et al. (2000) The shape of 433 Eros from the NEAR-Shoemaker Laser Rangefinder. *Science* 289: 2097–2101.
- Zuber MT, Smith DE, Lemoine FG, and Neumann GA (1994) The shape and internal structure of the Moon from the Clementine mission. *Science* 266: 1839–1843.
- Zuber MT, Solomon SC, Phillips RJ, et al. (2000a) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

10.06 Exogenic Dynamics, Cratering and Surface Ages

B. A. Ivanov, Institute for Dynamics of Geospheres, Moscow, Russia

W. K. Hartmann, Planetary Science Institute, Tucson, AZ, USA

© 2007 Elsevier B.V. All rights reserved.

10.06.1	Introduction	207
10.06.2	Impact Craters – Morphology	209
10.06.3	Basics of Impact Cratering Processes	211
10.06.3.1	Principal Stages of Impact Crater Formation	211
10.06.3.1.1	Contact and compression – shock wave generation	211
10.06.3.1.2	Excavation and transient crater growth	211
10.06.3.1.3	Transient cavity modification – slumping or collapse?	211
10.06.3.2	Impact Crater Scaling Laws	211
10.06.3.2.1	Impact velocity scaling	212
10.06.3.2.2	Strength and gravity scaling	213
10.06.3.2.3	Impact angle scaling	215
10.06.3.2.4	Impact heating, melting, vaporization	216
10.06.4	Size-Frequency Distribution of Impact Craters	217
10.06.4.1	Hartmann Production Function	218
10.06.4.2	Neukum Production Function	219
10.06.5	Impact Rate Estimates	221
10.06.5.1	Gravitational Mechanisms	222
10.06.5.2	Nongravity Mechanisms	222
10.06.5.3	Impact Probability	222
10.06.6	Extraterrestrial Surface Dating	225
10.06.6.1	Principles of Surface Dating by Crater Counting	225
10.06.6.2	Earth–Moon System	228
10.06.6.3	Mercury	231
10.06.6.4	Venus	232
10.06.6.5	Mars	232
10.06.6.6	Problem of Interpretation of Small Craters	232
10.06.6.6.1	Secondary craters on Mars	234
10.06.6.6.2	SFD for secondary craters	235
10.06.7	Conclusions	236
References		236

10.06.1 Introduction

Exogenic dynamics of planets (in contrast to endogenic dynamics) refers to the exchange of solid material among different zones of the solar system. Due to the gravity of planets and mechanics of rotation around the Sun, this material reaches the solid surface of a targeted planet with a high speed (from a few km s^{-1} to $\sim 100 \text{ km s}^{-1}$). The collision with the planetary solid surface (or with a planetary atmosphere) decelerates the falling body (the ‘projectile’) converting its kinetic energy into heat, light radiation, shock waves, and creating cavities – impact

craters – at the surface of targeted planets. High-velocity impacts redistribute ejected material, create a temporary ‘hot spot’ in the crust, melt, fracture, remagnetize target rocks, displace them, and sometimes, eject a small part of target rocks out of the target planetary gravity field. These ejecta can land at another planetary body, providing the material exchange between planets. Currently, we find Martian (SNC) and lunar ‘meteorites’, as well as asteroidal fragments on Earth. (SNC is the abbreviation for Martian meteorites historically (before recognition as rocks from Mars) named Shergottites-Nakhrites-Chassignites.)

The efficiency of the exogenic dynamics and modification rate of planetary surfaces depends on the frequency of impacts of a given energy and the geological evolution of the impact site (gravity readjustment of impact craters, cooling of impact ‘hot spots’). The frequency of impacts, in chain, depends on the ‘strength’ of small body sources, and the efficiency of delivery to the target planet. Both parameters may evolve, being different in the first 0.5 Ga after the solar system formation (ancient bombardment rate), and today (the modern bombardment rate). The geological evolution of the impact site depends on the planetary mechanical and thermal structure and the velocity of impact. As the main differentiation of terrestrial planets occurred relatively soon after the planet’s formation, the main difference in impact reaction during ancient bombardment now results from generally hotter interiors of ancient planets.

The projectile’s source strength is intimately connected to the history of the solar system formation. The current paradigm of the solar system formation assumes the collapse of the protosolar dust-gas nebula with the ignition of the star (Sun) in the center. Following events included dust, dust/ice, and ice accretion in sizable bodies during the continuing gas removal by solar radiation as discussed in pioneering work by (Safronov, 1969; Wetherill, 1978, 1980). The dust/ice ratio in growing bodies was controlled by the solar insolation, limiting the presence of liquid volatiles close to the Sun by heating. This resulted in formation of protoplanetary bodies with smaller volatile content within the modern Jovian orbit with semimajor axis $a = 5.2 \text{ AU}$ (1 AU – astronomical unit – is equal to the Earth–Sun distance of $\sim 1.5 \times 10^6 \text{ km}$), and protoplanetary bodies with much larger volatile (frozen gases) content beyond that. The feeding zone in the central part of the solar system produced terrestrial planets (Mercury, Venus, Earth, and Mars), their satellites (the Earth’s Moon is the main one), and a swarm of small bodies – asteroids – in the middle between orbits of Mars and Jupiter – by mutual collision/accumulation events of pre-planetary small bodies – planetesimals. Jupiter and other giant planets (Saturn, Neptune, Uranus) accumulated a huge amount of gases before the total dissipation of gaseous nebula. After the Sun, Jupiter is the next most massive and gravitationally important body in the solar system; hence, it set up resonance zones that affect motion of other planets, small bodies, and their fragments (see Chapter 10.13 for a discussion on the giant planets).

The early stage of the planetary formation is controlled by celestial mechanics of small bodies revolving

around the Sun with orbit change due to mutual collisions and (possibly) the drag of gaseous nebula residuals. The growth of a few main planets results in ‘excitation’ of orbits of smaller bodies and, possibly, in mixing of individual planetary feeding zones at the end of main planet formation. The timescale of terrestrial planet formation is $\sim 100 \text{ Ma}$. In any given zone more than one large planetesimal may have emerged, and it is very probable that the terrestrial planet formation included a few giant impacts of protoplanets with diameters of 1000 km as discussed by Hartmann and Davis (1975) in early work on lunar origin.

Beyond Mars the presence of growing Jupiter exited the orbital motion of local small protoplanetary bodies (the origin of planetary satellites is discussed in Chapter 10.14). Collision velocities increased, fragmentation began to dominate over accretion, and this limited size of the largest bodies, accumulated in this zone. Currently, the largest body here is Ceres with diameter of $\sim 1000 \text{ km}$. One cannot exclude that Mars-size bodies could have originally been accumulated here or elsewhere among terrestrial planets, removed soon by gravity action of Jupiter and mutual gravity interaction between largest protoplanets (Chambers and Wetherill, 2001). The flux of these bodies together with ‘leftovers’ from terrestrial planet’s feeding zones created the early heavy bombardment of just-formed crusts on terrestrial planets. The remnant of this zone is now the main belt (MB) of asteroids. The largest MB asteroids ($>\sim 100 \text{ km}$ in diameter) are most probably leftovers of the primary population, smaller ones are definitely fragments, formed during the collision evolution. A number of small bodies may change their MB orbits into planetary-crossing orbits (PCOs) under the action of resonances with Jupiter and Saturn. The motion of PCO bodies also evolve due to weak resonances with Mars and close encounters with terrestrial planets. The small body lifetime at the typical PCO is between 1 and 30 Ma (Gladman *et al.*, 2000), and a number of ‘generations’ of bodies, supplied from MB, should appear before one of them strikes the planetary surface.

The planet accumulation rate is strongly controlled by the period of revolution around the Sun. At the distance of Earth the body makes the full revolution during 1 year. At the distance of Jupiter, for ~ 11 years; at the distance of modern Uranus, the revolution period is as large as ~ 84 years; and Neptune, ~ 165 years. The theory of fast accumulation, applicable for terrestrial planets, would give too large time of formation of Uranus in the observed position. Hence, the set of hypothesis has been proposed to permit Uranus and

Neptune or the formation of their embryos (Zharkov and Kozenko, 1990) closer to the Sun with following migration to the modern orbits (Gomes *et al.*, 2005; Tsiganis *et al.*, 2005). The migration of giant planets, as well as their formation (mass concentration appearance), would ‘excite’ orbits of small bodies, producing a ‘bombardment epoch’ for inner planets and ejecting a part of small bodies far from the Sun. These bodies – ‘leftovers’ of giant planet formation – shape the periphery of the solar system: the spherical shell of comets (cometary nuclei, in a strict sense) and the Kuiper–Edgeworth object’s (KBO) belt beyond the Neptune orbit. The Pluto–Charon binary system, had it been discovered today, would have been cataloged as a member of KBO (Bertoldi *et al.*, 2006; Sheppard 2006). In 2006 Pluto’s official status was changed from ‘planet’ to ‘dwarf planet’ (Green, 2006) (see Chapter 10.01 for definitions of planets and dwarf planets).

Occasional returning of comet nucleus and KBOs (KBOs are discussed in Chapter 10.16) close to the Sun creates the comet bombardment population of small bodies. Comets, spending appreciable time close to the Sun, may lose most of volatiles, creating the subpopulation of dormant/extinct comets, occupying asteroid-like orbits. The share of active/dormant/extinct comets in the total bombardment flux on terrestrial planets is a matter of current debates (Fernandez *et al.*, 2002; Stuart and Binzel, 2004). Authors are in favor of a minor role (<15%) of comet impacts for terrestrial planets (Ivanov *et al.*, 2002). For satellites of giant planets, comets seem to be the main bombardment population (Zahnle *et al.*, 1998, 2003). Estimates of absolute impact fluxes in the outer solar system, however, is an extremely difficult problem at present levels of knowledge, since interplanetary source populations are poorly known. (The formation of the satellites of the giant planets is discussed in Chapter 10.14.)

The relative role of various projectile populations is still under investigation. The author’s point of view (subjected to further work) has been published in a set of papers (Hartmann, 1977; Ivanov, 2001, 2004; Ivanov *et al.*, 2001, 2002; Hartmann *et al.*, 2002).

Authors are unable to cover here the very important topic of cratering on icy satellites of the outer planets. At some level, the whole system of estimating impact fluxes and absolute ages on terrestrial planets breaks down for the outer solar system. The planet crossing bodies that might hit Jupiter and Saturn satellites, for example, are so distant that statistics are very poor regarding the size-frequency distribution (SFD) at small sizes. The dynamical methods are compromised because there may be various poorly studied resonant

effects throwing objects onto PCOs (as with near-Earth objects (NEOs) from MB resonances), but the parent populations such as KBOs, Centaurs, and Trojan swarms are probably not fully characterized. Surfaces such as Europa, Titan, and parts of Enceladus, with notably small crater densities are believed to have geologically young absolute ages, indicating ongoing heating and geological activity, but we believe the error bars on ages must be much larger than the factors 2–4 that have been quoted for Mars. The most complete discussion of the impact cratering and cratering rate on outer planet satellites is presented elsewhere (Zahnle *et al.*, 1998, 2003; Schenck *et al.*, 2004).

10.06.2 Impact Craters – Morphology

Impact craters are circular depressions with elevated rims (Figure 1) created by impacts of celestial bodies. The impact crater formation occurs as a chain of subsequent processes, started with a high-velocity impact and ending with ejected material deposition and crater cavity adjustment to the new equilibrium of fragmented rocks in the planetary gravity field. First observations of impact craters have been done with telescopic observations of the Earth’s Moon starting with Galileo. Impact craters were eventually recognized on Earth’s surface, particularly in the 1950s and 1960s (Dence, 1965; French and Short, 1968). With the beginning of space flights the scientific interest has grown enormously, and following progress occurred in parallel both with imaging of extraterrestrial planetary bodies and with studying of terrestrial impact craters. The close-up imaging of the Moon has revealed the fundamental difference of the impact crater population

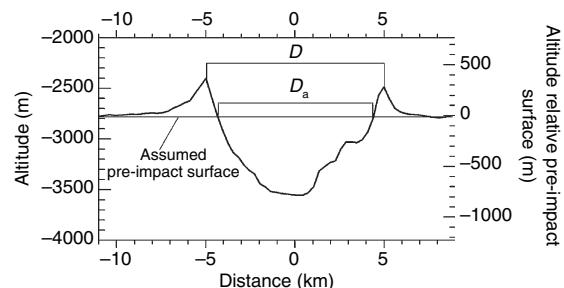


Figure 1 The profile of a typical fresh impact crater (exemplified with Martian Zunil Crater, profiled by MOLA track). North-to-south profile does not reveal the small central mound visible in high-resolution images. The figure introduces the often used rim crater diameter, D , and the apparent crater diameter, D_a , measured at the pre-impact surface level.

from other (e.g., volcanic) crater structures, namely – the wide spectrum of crater sizes, continuing from microns to thousands of kilometers.

The size-morphology progression ranges from small, simple bowl-shaped craters, through craters with central peaks and terraces at the walls, up to basins with inner rings and, on some bodies, exterior scarps that encircle the impact site.

In terms of understanding planetary geology we are fortunate that nature ‘stamps’ sharp, symmetric circular structures on planetary surfaces. If the crater formation rate is understood, we can not only estimate ages from crater numbers, but also interpret nature and rates of geologic erosion and deposition processes from morphological degradation and losses of small craters.

It is currently believed that this morphologic diversity is not a direct result of the crater excavation process, but develops only as the explosion cavity evolves just after most of the material has been ejected from the initial cavity. The initial product of crater excavation is believed to be a circular, bowl-shaped ‘transient’ cavity with a depth/diameter ratio between 1/4 and 1/3. This ‘transient’ crater then undergoes different degrees of modification as a result of gravitational instability and collapse. The final crater morphology depends on the acceleration of gravity, density, and condition of the surface materials in the target planet. This affords both the opportunity of using impact crater morphology to learn about conditions in the target, as well as the challenge of understanding how different conditions affect the course of crater collapse (Melosh and Ivanov, 1999).

M. R. Dence (1965) first classified terrestrial impact craters as either simple or complex in structure. This classification, based on the study of craters exposed by erosion, also seems to apply to the morphology of fresh craters revealed by images of the surfaces of other planets and satellites. Data sets on terrestrial and planetary craters are highly complementary: due to erosion, geophysical investigations, and direct drilling, the subsurface structure of many terrestrial craters can be explored in detail. The surface morphology is often displayed in exquisite detail by images of fresh craters on the surfaces of other planets or satellites.

Simple craters are circular, bowl-shaped depressions with raised rims and approximately parabolic interior profiles. The rim-to-floor depths of a large sample of simple craters on the Moon are roughly 1/5 of the rim-to-rim diameter (Pike, 1977a). Simple

craters have a true floor, covered with a breccia lens up to a half of the ‘true crater’ depth. Drilling results and seismic investigations prove that this lens is created by fragmented material avalanched from the transient crater walls down to the crater, forming the ‘apparent’ crater floor (Shoemaker, 1963; Grieve *et al.*, 1977). The volume of the breccia lens is roughly one-half the volume of the crater itself (Settle, 1980; Grieve and Garvin, 1984).

Complex craters possess a more complicated structure than simple craters. On the Earth they exhibit central structural uplifts, rim synclines, and outer concentric zones of mainly normal faulting. Images from spacecraft show extraterrestrial craters with single or multiple central peaks, flat inner floors, and terraced rims. The depths of complex craters increase with increasing diameter, but much more slowly than the depths of simple craters (Pike, 1977a, 1977b).

The transition between simple and complex craters occurs over a relatively narrow diameter range on any given solar system body and seems to scale as the inverse power of the surface gravity, (e.g., Hartmann, 1972). Thus, the simple–complex transition (Pike, 1980) is relatively well determined on the Moon at about 15 km diameter. On Mercury and Mars it is about 11 and 7 km diameter correspondingly, and 3–4 km diameter on Earth (depending on whether the crater forms in sedimentary or crystalline rocks). On icy satellites the transition depends on the low-temperature ice strength and the presence or absence of a subice liquid mantle (Moore *et al.*, 2001; Schenk, 2002; Schenk and Ridolfi, 2002; Schenk *et al.*, 2004).

Hartmann and Wood (1971) described a broadening of central peaks into ring of peaks, for which they coined the term ‘peak-ring’. Peak-rings may mark a transition to full-fledged multi-ring basins.

Impact basins are largest impact structures; on some bodies they exhibit many rings, being named as multi-ring basins. The classic basin is Orientale on the Moon’s western limb (eastern limb in terms of sky coordinates used by early telescopic observers). First recognized on rectified lunar photographs in 1962 (Hartmann and Kuiper, 1962), multiring basins appear to form as a tectonic response of the target’s lithosphere to the crater created by the impact (Melosh and McKinnon, 1978; McKinnon and Melosh 1980).

A review of morphology and occurrence of basins is published by Spudis (1993). The data from recent space planetary missions resulted in the

enlarged volume of data about basins and their deep structure on the Moon (Pieters *et al.*, 1997; Wieczorek and Phillips, 1999; Potts and von Frese, 2003a, 2003b; Sugano and Heki, 2004; Crosby and McKenzie, 2005), and Mars (Zuber *et al.*, 2000; Hiesinger and Head III, 2002; Neumann *et al.*, 2004; Potts *et al.*, 2004). New data about Mercury have also been published (Andre *et al.*, 2005; Andre and Watters, 2006).

10.06.3 Basics of Impact Cratering Processes

10.06.3.1 Principal Stages of Impact Crater Formation

The impact process as a whole can be described as a kind of explosion, in which the initial kinetic energy of the projectile does work on the target to create a visible depression – the crater – as well as to heat both projectile and target. The sequence of events in a high-velocity impact may be presented as a set of several sequential stages. Each of these stages is dominated by a specific set of major physical and mechanical processes. It is conventional to distinguish three main stages of an impact event (Melosh, 1989). These stages are contact and shock compression, transient cavity growth by crater material ejection, and finally, transient cavity modification (slumping or collapse). These stages do not have strict boundaries and are used for convenience in the analysis of impact processes because they highlight the dominant mechanisms acting at any given time.

10.06.3.1.1 Contact and compression – shock wave generation

In the first stage of an impact, the energy released by deceleration of the projectile results in the formation and propagation of shock waves away from the point of impact. During this stage shock pressures are of the order of the stagnation pressure ρU^2 , where ρ is the smaller of either the projectile or target density, and U is the impactor velocity. At speeds above a few km s^{-1} these pressures greatly exceed the strengths of rocks (a few GPa at high pressure) and the process can be treated hydrodynamically. Shock waves decay propagating out of the point of impact transform to stress waves, and at larger distances, to the elastic (or seismic) waves.

10.06.3.1.2 Excavation and transient crater growth

After the shock wave expands away from the impact site, accelerated target materials move out of the impact point, opening the transient cavity of a growing impact crater. Studies of the crustal thickness beneath large lunar basins also support this hypothesis of ‘proportional scaling’ for the dimensions of the transient crater up to craters at least 500 km in diameter (Wieczorek and Phillips, 1999). In many numerical model simulations, the floor of the transient crater achieves its maximum depth and begins to rise into a central peak while the diameter is still increasing (Melosh, 1989).

10.06.3.1.3 Transient cavity modification – slumping or collapse?

Details of the transient cavity modification by wall slumping for simple craters are presented by Grieve and Garvin (1984). The full-scale review of mechanics of transient cavity collapse in the case of complex crater formation is given by Melosh and Ivanov (1999). General review of complex crater formation presented by Melosh (1989) is still the best introduction into the problem of complex crater formation. Some particular issues are discussed in recent papers about numerical modeling of impact craters (O’Keefe and Ahrens, 1993, 1999; Morgan *et al.*, 2000; O’Keefe *et al.*, 2001; Collins *et al.*, 2002; Ivanov, 2005).

10.06.3.2 Impact Crater Scaling Laws

Scaling laws relevant to impact cratering may be defined as a set of rules (experimentally derived or theoretically assumed), formulated to establish relationships between parameters of an impactor (projectile) and the resulting impact crater under conditions of different planetary bodies. Originally scaling laws were formulated to make a theoretical interpolation of laboratory experimental data, mostly for metal targets (Kinslow, 1970), and to extrapolate lab-scale experimental data to natural impact velocity range. Later rock/soil targets attracted attention as planetary scientists tried to estimate the size of a projectile that created a crater of interest. Rock/soil targets were also in the focus of explosion cratering studies (Roddy *et al.*, 1977). For metal targets, the shear strength of metals plays the main role to limit the size of the growing transient cavity, but for natural planetary targets, where crater diameter ranges from a few meters to >1000 km, one observes a

transition from strength-dominated formation of small craters to gravity-dominated formation of large craters. ‘Large’ and ‘small’ crater diameters depend on intrinsic target strength, which may change in a complicated way during the cratering process (brittle damage of intact rocks, additional fragmentation of real fractured rocks, mechanical dispersion of shock energy in sands and regoliths, temporary decrease of friction, thermal softening – vanishing of rock strength near to melting point). For this reason, it is convenient to refer to an ‘effective’ rock strength, which can be used as a parameter in available theoretical models.

Below, we review main approaches to scale cratering effects for a variety of impact velocities, target strength parameters, and an impact angle. It is important to compare crater sizes for same-size projectiles on different planets. The reliability of this comparison defines the accuracy of interplanetary crater retention age recalculations.

Early theoretical study (Zeldovich, 1956) has resulted in the concept, later named as ‘the late stage equivalence’ (Dienes and Walsh, 1970), formulated as a similarity of mechanical action, measured at distances much larger than the projectile size, for all impacts in the same target is the same, if the value of a so-called ‘coupling parameter’, is constant.

10.06.3.2.1 Impact velocity scaling

Dienes and Walsh (1970) have proposed a coupling parameter to scale the shock pressure decay under the point of impact in the form of a characteristic length scale, L , for a given impact:

$$L = D_p(\delta/\rho)^v [U c_b]^\mu \quad [1]$$

where D_p is the projectile diameter, δ/ρ is the projectile/target density ratio, U is the impact velocity, and c_b is the bulk velocity in a target material, defined as $c_b = (K/\rho)^{0.5}$, where K is the bulk compression modulus, v and μ are dimensionless exponents, derived from experiments.

The central point of the ‘coupling parameter’ concept is that the cratering efficiency of an impact will be the same for all projectiles with the same coupling parameter value. For the density ratio exponent the value of $\zeta = 1/3$ gives a good approximation for available data. For the same projectile density a simple ‘energy’ scaling demands that the value $D_p v^{2/3} = \text{const.}$, while experiments and numerical modeling reveal the ‘equivalent mechanical action’ (e.g., the same crater dimensions) for projectiles

with $D_p v^\mu = \text{const.}$ and the velocity exponent $\mu < 2/3$ ($\mu \approx 0.55–0.58$ for competent nonporous targets and $\mu \approx 0.4$ for porous targets like dry sand). The fundamental reason for the inequality $\mu < 2/3$ is that at higher impact velocity a small amount of material near the impact point is overheated and ejected decreasing the amount of energy, available for the crater formation at later stages. The lower theoretical limit (‘momentum scaling’) demands that $\mu > 1/3$ (Holsapple and Schmidt, 1982; Holsapple, 1993).

The scaling of depth, z , below the impact point with the L value allows us to present the shock pressure decay in the form

$$\frac{p}{\rho c_b^2} = f_p(z/L) \quad [2]$$

where f_p is the universal function for a wide range of impact conditions. Experimental data for metal targets may be found in Dienes and Walsh (1970). The progress in the numerical modeling resulted in the investigation of a shock pressure decay in geomaterials such as rocks. Barely all rocks, typical for planetary crusts, demonstrate the presence of a solid–solid phase transition under compression (quarts–coesite–stishovite transition is a typical example). This makes scaling not so simple as it is for metal targets. However, available numerically modeled data show reasonably good converging to a general shock pressure decay function being presented in a dimensionless form of eqn [1]. For target material with simple equations of state, the shock pressure decay may be expressed as (Ivanov, 1981)

$$\frac{p}{\rho c_b^2} = (3 \pm 1) \times (z/L)^{-2.5} \quad [3]$$

with the best choice of $\mu = 0.58$, recommended by Dienes and Walsh (1970).

Pressure decay in geomaterials with the solid–solid phase transition and with initial porosity, created the dissipative processes for the sink of energy into an additional residual heating of the shocked material, showing more steep decay with a distance from the impact point.

Holsapple and Schmidt (1980, 1982, 1987) and Schmidt and Housen (1987) have found a set of parameters to scale cratering dimensions without explicit usage of target material compressibility. Despite this set of parameters not guarantying the similarity of shock waves, it has been proved to be useful in many applications of ‘high velocity’ impacts when impact velocity is larger than the sound speed in the target and the projectile is not infinitely strong.

For ‘low-velocity’ impacts (subsonic relative to the target material sound speed and/or deeply penetrated projectile), the coupling parameter function may be different, particularly if it depends on impact velocity not as an exponential function. For example, low-velocity impacts in sand are well described with a coupling parameter proportional to a linear combination of v and v^2 terms (Mizutani *et al.*, 1983). Below, we concentrate on high-velocity impacts.

10.06.3.2.2 Strength and gravity scaling

For the given value of the coupling parameter, the kinetic energy of the excavation flow may be spent mostly as a plastic work against strength of target material (‘strength regime’), or mostly as the potential energy of the growing cavity in the gravity field (‘gravity regime’). In the former case the energy consumed during the transient cavity growth is proportional to $Y\alpha^3$ (the multiplication of a transient crater volume, proportional to α^3 by the effective strength Y), while in the latter case the consumed energy is proportional to ρga^4 (the multiplication of the mass extruded from the transient cavity, proportional to $\rho\alpha^3$, by the gravity acceleration, g , and the effective height of the mass uplift, proportional to a). The gravity craters need more and more energy to make a larger crater. It results in large craters which are smaller in respect to the projectile and shock melt zone size. Largest impact basins on terrestrial planets may have transient cavity only a few times larger than the projectile size (exact proportions depend on the assumed impact velocity). In the intermediate crater size range both gravity and strength energy consumption mechanisms operate simultaneously.

There are many ways to express the transition from strength to gravity regimes. The basic experimentally derived scaling laws (Schmidt and Housen, 1987) use π -theorem (see details in Melosh (1989)), and in the original formulations may be presented as follows:

$$\pi_D = K\pi_2^{-\alpha} \quad [4]$$

where π_D is the scaled crater diameter,

$$\pi_D = D_a(\rho/m)^{1/3} \quad [5]$$

The scaled projectile diameter (for an equivalent in mass sphere of D_p diameter) in the gravity-dominated regime is

$$\pi_2 = (1.61gD_p)/U^2 \quad [6]$$

and, in the strength-dominated regime, is

$$\pi_2 = Y[\rho U^2] \quad [7]$$

where Y is an effective strength of the target material.

Here D_a is the transient crater apparent diameter, ρ is the target density, m is the projectile mass, U is the impact velocity, g is the gravity acceleration, and D_p is the projectile diameter. The processing of experimental data (Schmidt and Housen, 1987) results in determination of coefficients in eqn [4] for various materials. The most useful rock types are as follows: for nonporous rocks $K=1.6$ and $\alpha=0.22$ and for porous (dry sand) targets $K=1.68$, $\alpha=0.17$

Figures 2(a) and 2(b) illustrate the scaling law for dry sand and nonporous rocks in comparison with most of the experimental data used to define scaling constants. For porous rocks (dry sand, lunar regolith) the scaling law looks well justified (**Figure 2(a)**). For nonporous rocks the wet sand data have been used as a proxy. Schmidt and Housen (1987) specially outline that crater size values for wet sand target the pure exponential regime is not achieved in their experiments, being the justified extrapolation to the pure gravity scaling from experimentally covered strength/gravity transition region. The experimental laboratory data may be expanded with the numerical modeling results, where both transient crater diameter and the final rim diameter have been recorded. In **Figure 2(b)** numerical model data for terrestrial craters with rim diameters from ~ 5 to ~ 200 km in the layered targets (sediments over granitic crust over dunite mantle) looks as a natural continuation of high-G experimental data, keeping however, less steep slope in $\pi_D-\pi_2$ coordinates (close to a slope for porous targets). The reason for this behavior may be the presence of dry friction in numerical models, not well discussed in early nonporous scaling law derivation (Schmidt and Housen, 1987). Both experimental (Schmidt, 1980) and numerical modeling (Wünnemann *et al.*, 2006) data show that porosity and dry friction in targets should be decoupled in an updated version of scaling technique now under construction.

The π -theorem relations may be rewritten in the following familiar form: for nonporous rocks

$$D_{tc} = 1.16(\delta/\rho)^{1/3} D_p^{0.78} (U)^{0.44} g^{-0.22} \quad [8]$$

and For dry sand

$$D_{tc} = 1.25(\delta/\rho)^{1/3} D_p^{0.83} (U)^{0.34} g^{-0.17} \quad [9]$$

Relations presented above are valid for the gravity regime of cratering. The procedure of approximate

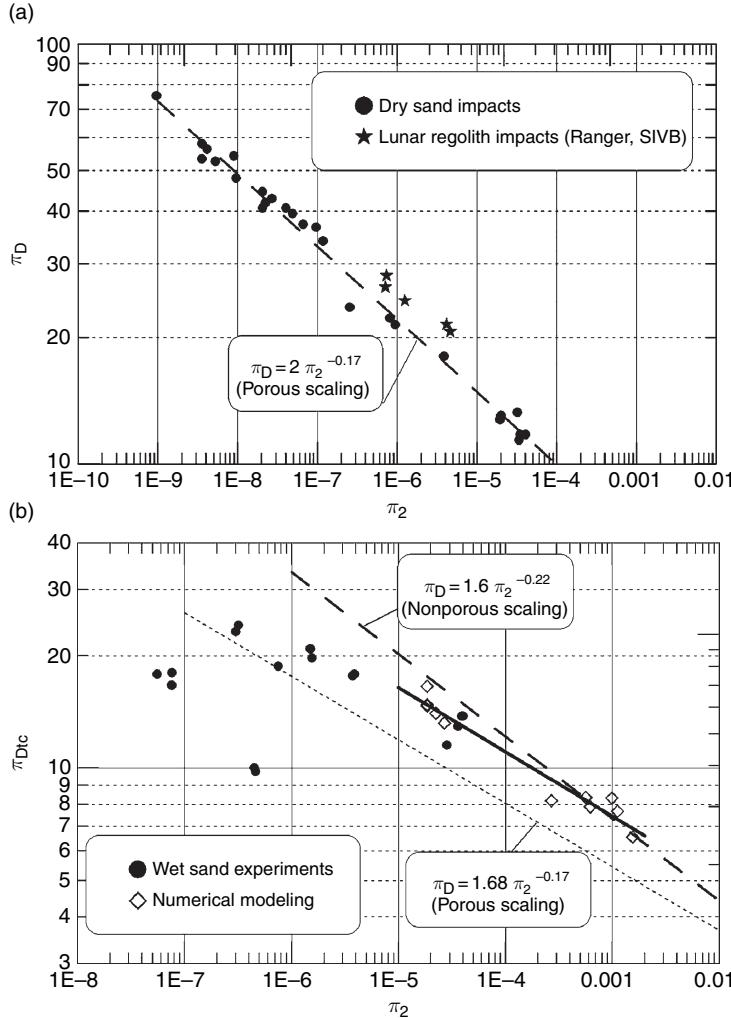


Figure 2 (a) π -Scaling for dry sand ('porous rock') for the rim crater diameter. Lab experimental points (Schmidt, 1987; Housen and Holsapple, 2003) are compared with artificial lunar craters created by Ranger and SIVB spacecraft impacts (Moore, 1972; Witaker, 1972). (b) Scaled transient crater radius π_{Rtc} vs scaled projectile size, π_2 , from experimental Boeing Lab data (Schmidt and Housen, 1987) for wet sand (black circles), and from numerical modeling (Stöffler *et al.*, 2004; Ivanov, 2005; Ivanov and Stöffler, 2005; Joeleht *et al.*, 2005) as open diamonds, transient cavity is defined as in (Wünnemann *et al.*, 2006). Numerical modeling data looks like the natural continuation of experimental data, following however the exponent -0.17 , similar to experimental exponent for dry sand but not as steep as assumed 'nonporous' value of -0.22 (Schmidt and Housen, 1987). Is it the indication of a finite friction in the numerical model?

including of the strength–gravity transition, inspired by Holsapple and Schmidt (1979), is described in Neukum and Ivanov (1994) and Ivanov *et al.* (2001). Arithmetically, the engineering equation is transformed to

$$(D_{tc}/D_p)^{0.78} = 1.16(\delta/\rho)^{1/3}(U)^{0.44}(gD_{tc})^{-0.22} \quad [10]$$

and, consequently, to

$$D_{tc}/D_p = \left[1.16(\delta/\rho)^{1/3}(U)^{0.44}(gD_{tc})^{-0.22} \right]^{1/0.78} \quad [11]$$

Finally, combining all numerical values for assumed δ , ρ , U , into one coefficient C_D ,

$$D_{tc}/D_p = C_D/(gD_{tc})^{0.282} \quad [12]$$

In the strength regime the ratio of D_{tc}/D_p is a constant. To make a smooth transition from the strength to gravity regime we assume that

$$D_{tc}/D_p = C_D/[g(D_{sg} + D_{tc})]^{0.282} \quad [13]$$

where D_{sg} is the effective boundary between gravity- and strength-dominated regimes of a crater formation – see the discussion on the actual D_{sg} value for terrestrial rocks in Holsapple and Schmidt (1979), Croft (1981), and Neukum and Ivanov (1994).

Here we have the smooth approximation of the transition to strength craters: for small craters ($D_{sg} \ll D_{tc}$), the D_{tc}/D_p ratio is constant for the given CD value:

$$D_{tc}/D_p = C_D/[gD_{sg}]^{0.282} \quad [14]$$

The value of $1/[gD_{sg}]^{0.282}$ plays the role of the effective target strength parameter (smaller the strength, larger is the crater, and smaller is the transitional diameter between craters formed in strength and gravity regimes).

Transient cavity diameter is proportional to rim crater diameter as $D \approx (1.25 \pm 0.05)D_{tc}$ for simple craters, and, according to Croft's model (Croft, 1985)

$$D_{tc} = D_{sc}^{0.15} D^{0.85} \quad [15a]$$

or, vice versa,

$$D = D_{sc}^{-0.18} D_{tc}^{1.18} \quad [15b]$$

for $D > D_{sc}$, where D_{sc} is the simple-to-complex crater transition diameter. One should note that the simple-to-complex transition is poorly investigated and the recipe mentioned above is the simplest (but internally consistent) one. Equation [15] was obtained from geologic reconstruction of partially eroded terrestrial craters. Numerical modeling gives an opportunity to get similar relations directly by comparison of transient cavity diameter and the final rim diameter of a collapsed crater (Figure 3). The exponent in eqn [15b] is close to model value derived in Figure 3 proving the general validity of Croft's model but opening room for further natural and model data reconciliation.

For other planetary bodies the D_{sc} value may be assumed from crater morphology transition (Pike, 1980).

The numerical modeling of the terrestrial impact crater formation, extended to the stopping of the target material motion (see references in the caption for Figure 2), allows us to write down the approximate relationship between projectile diameters, D_p (km), impact velocities, U (km s⁻¹), and the final rim crater diameter, D (km) for the vertical impact as

$$D = 4(D_p U^{0.58})^{0.91} \quad [16]$$

what is valid for the diameter range shown in Figure 3.

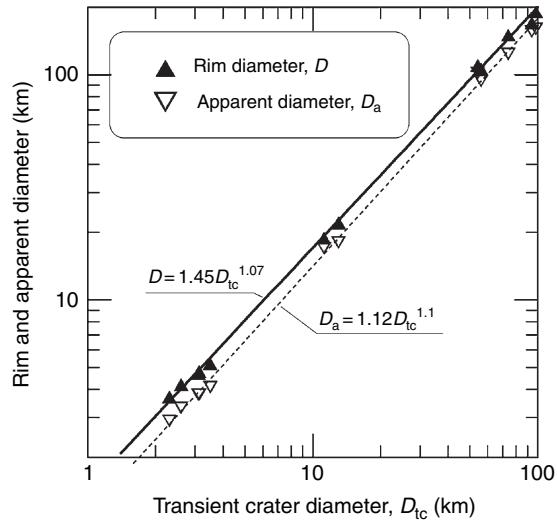


Figure 3 The relation between the transient crater diameter vs rim crest and apparent final complex crater diameters from a direct numerical modeling of terrestrial craters with diameters from ~4 to ~200 km (see references in Figure 2(b)).

The general behavior of D/D_p ratio for craters in lunar nonporous rocks has been presented in Neukum and Ivanov (1994) and (Ivanov *et al.* (2001)). The general D/D_p dependence on final crater diameter for lunar craters is shown in Figure 4, constructed under assumptions that the impact velocity is 18 km s⁻¹, lunar regolith density is 1900 kg m⁻³, projectile and bedrock density is ~2700 kg m⁻³. Later in the text eqn [16] and scaling laws shown in Figure 4 will be used to compare size-frequency distribution of projectiles and impact craters on Earth and the Moon (Section 10.06.6.2).

10.06.3.2.3 Impact angle scaling

Vertical impacts of celestial bodies into planetary surfaces, suitable for computational and experimental study, have a low probability (as low as the probability of grazing impacts). As it was shown for nongravity spheres (Gilbert, 1893) and for spherical planets with the natural gravity attraction (Shoemaker, 1962), the probability of impacts of a random flux of projectiles from infinity at various angles has a ‘domical’ angle distribution with the maximum at 45°:

$$f_\phi = 2 \sin \phi \cos \phi d\phi \quad [17]$$

where f_ϕ is the probability of impact in the interval of angles from ϕ to $\phi + d\phi$. Here we agree to measure angle of impact from the horizon ($\phi = 0^\circ$ for grazing impacts, $\phi = 90^\circ$ for vertical impacts). It is believed

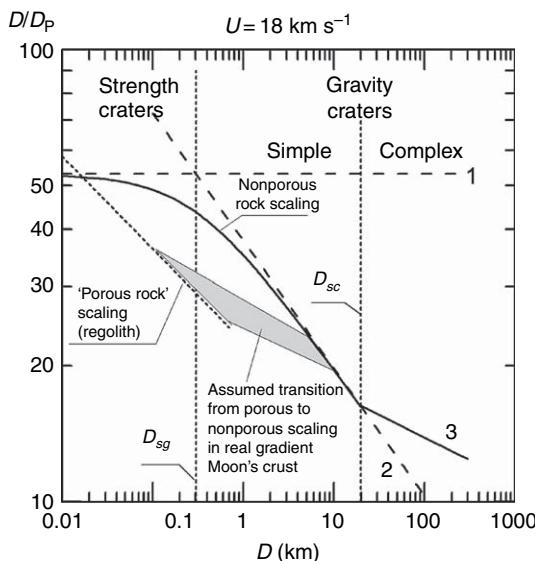


Figure 4 An example of the lunar crater scaling. Upper layer of the Moon is constructed of porous regolith and for the assumed 18 km s^{-1} impact velocity the crater/projectile diameter ratio is comparable with the strength scaling for craters in nonporous bedrocks. For craters in nonporous bedrocks the D/D_p ratio gently declines from strength scaling ($D/D_p = \text{const}$, dashed line 1) toward the gravity scaling (dashed line 2), but at $D \sim 15 \text{ km}$ craters undergo widening due to the gravity collapse. For the ‘real’ layered lunar crust with a gradual transition for a relatively porous megaregolith to the solid rocks one may assume a range of transition from the ‘porous’ to ‘nonporous’ scaling for D between ~ 0.5 and $\sim 10 \text{ km}$.

that eqn [17] is valid for planets and projectiles revolving around the Sun.

The oblique impact is less efficient than the vertical impact. The experimental modeling reveals, for example, that for the same projectile and the same impact velocity the volume of crater, V , decreases approximately as

$$V(\phi) = V(90^\circ) <1> \sin^\gamma \phi \quad [18]$$

where $V(90^\circ)$ is the crater volume for vertical impacts and the exponent γ is in the range from 1 to 2 depending on the target properties. Gault and Wedekind (1978) propose γ between 1 and 2 for impacts in monolithic granite blocks (aluminum projectiles), and $\gamma=1$ for impact in porous targets prepared of noncohesive dry sand and volcanic adhesive pumice. These results are widely cited despite the obvious difference in impact conditions – at the maximum impact velocity, v , of 7 km s^{-1} impacts into sand and pumice are ‘supersonic’ (v is 4–5 times larger than the sonic speed in the target material) while for solid rocks it is ‘subsonic’ to ‘sonic’ regime

(v of $5\text{--}7 \text{ km s}^{-1}$ is close to the longitudinal elastic wave velocity). The data published by Gault and Wedekind (1978) may be approximately generalized assuming that the value of U in scaling laws for vertical impacts for an oblique impact may be replaced with $U \sin \phi$, presenting the vertical component of a projectile velocity (Pierazzo and Melosh, 1999, 2000). However, the nature of oblique impact looks more complicated. The repetition of such experiments for monolithic rocks (Burchell and Whitehorn, 2003) witnesses for validity of $\gamma=1$ similar to the porous target value obtained by Gault and Wedekind (1978). First last stage numerical modeling (Artemieva and Ivanov, 2001; Ivanov and Artemieva, 2001, 2002) revealed that for velocity of 20 km s^{-1} the efficiency of oblique impacts decreases with $\sin \phi$ no so dramatically as it should be according to ‘the vertical velocity component’ paradigm. The issue deserves further investigation. As a tentative recipe one can take into account that the 45° and steeper high-velocity impact (15 km s^{-1} and faster) may be as effective as the vertical impact.

10.06.3.2.4 Impact heating, melting, vaporization

The essential part of the initial projectile kinetic energy transforms into target material thermal energy, distributed around the crater. The further from the point of impact, the smaller is impact heating.

Close to the point of impact (in the zone or would be impact crater) irreversible shock wave heating (the residual internal energy after shock compression/decompression cycle) is large enough to push target material thermodynamic state above the boiling point. For typical impact velocity of 20 km s^{-1} and typical for terrestrial planets crust rocks (granite, basalt, anorthosite), the impact heating is not strong enough to make the total vaporization of rocks. The most probable here is the state of partial vaporization where rock melt at liquidus temperature (pressure dependent) is in equilibrium with gaseous phases, presented by various oxides with temperature-dependent chemistry (e.g., Yakovlev and Basilevsky, 1994). During the partial vapor cooling, the chemical content of liquid and vapor phases may vary with a possible quenching at about 2000 K. The following cooling and expansion may produce nonequilibrium products. Laboratory data about simulated rock vaporization are illustrated by Gerasimov *et al.* (1999).

For residual temperatures below the boiling point shock heating resulted in ‘impact melting’ of rocks. In

terrestrial craters the impact rocks are easily identified by texture. The important differences of impact melt in comparison with volcanic melts are: (1) the presence of large volumes of material melted above the solidus temperature (overheated melt) and (2) the presence of heavy fragment loading, untypical for volcanic melts. The history of fragments (clasts) incorporated in the impact melt may be very complicated – fine dust may be digested by melt, while large boulders may be ‘hanged’ inside the quickly solidified melt body.

With the exception of very large craters (diameters ~ 200 km and larger), the impact melt has a two-stage cooling history (Onorato, *et al.*, 1978). During the first stage of cooling the local thermal equilibrium with local clasts is achieved. Later the equilibrium ‘in gross’ (cooling) of the entire melt body happens by convection (in large craters) and heat conduction (in smaller craters, where local thermal equilibrium results in temperatures close to solidus and correspondingly large viscosity). In terrestrial craters with diameters below 100 km (like Manicouagan) clast cooling decreases impact melt temperature to solidus in about 100 s. On Venus larger crust temperature (~ 740 K) results in slower approaching to the local thermal equilibrium (due to smaller thermal gradient between clasts and melt) and increases the final melt temperature. Hence on Venus impact melt should be more mobile for ~ 10 times longer time (Ivanov *et al.*, 1992).

In large impact craters (like terrestrial crater Sudbury and larger), the volume of impact melt may be so large that the local thermalization finishes with temperatures above solidus and even liquidus open the possibility for a continuous magmatic process in the melt body, including convection (Zieg and Marsh, 2005).

The impact melt volume in a few terrestrial impact craters has been estimated (Grieve and Cintala, 1992; Cintala and Grieve, 1998a, 1998b), opening the possibility to compare theoretical values obtained from shock wave propagation and crater scaling laws. Today this is the only available method to verify scaling laws for large impact craters. This comparison made by several authors gives the estimated accuracy of a factor of 2 for the projectile mass (and, correspondingly, factor of $2^{1/3}$ for the projectile size) for assumed projectile density, impact angle, and velocity (Ivanov, 1981; Pierazzo *et al.*, 1997; Ugalde *et al.*, 2005). **Figure 5** illustrates the comparison between impact melt volumes observed and estimated with scaling laws for terrestrial impact craters in a wide range of crater sizes.

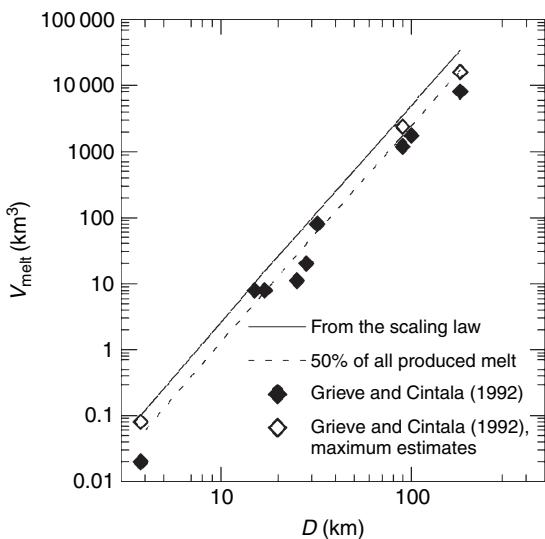


Figure 5 Comparison of estimates made using scaling laws (solid and dashed lines, constructed using data by Ivanov (1981) and Pierazzo *et al.* (1997)) and using geological mapping of terrestrial impact craters, summarized by Grieve and Cintala (1992).

10.06.4 Size-Frequency Distribution of Impact Craters

The Moon is an ideal test site to study cratering records, particularly since most lunar endogenic activity ended more than 3 Ga (gradually decaying in time (Hiesinger *et al.* (2000, 2003))). Thus, over the last 3 Gy or so, impacts alone have dominated modifications to the lunar landscape. Moreover, space missions have studied the Moon extensively (space mission orbital dynamics and instrumentation is discussed in Chapters 10.17 and 10.18, respectively). Returned samples of lunar landing sites provide a unique opportunity to ascribe an age to select craters and regions where accumulated impact craters have been counted as reviewed recently by Neukum *et al.* (2001a) and Stöffler and Ryder (2001). Hence, on the Moon we can estimate the cratering rate as the number of craters km^{-2} of a given diameter accumulated at a given surface during a given time interval.

The number of large craters is smaller than the number of small craters, reflecting the fact that the number of large projectiles in the solar system is smaller than the number of small projectiles. This fact has been recognized well before the active Moon exploration with spacecraft. **Figure 6** illustrates the early crater counts on lunar maria using

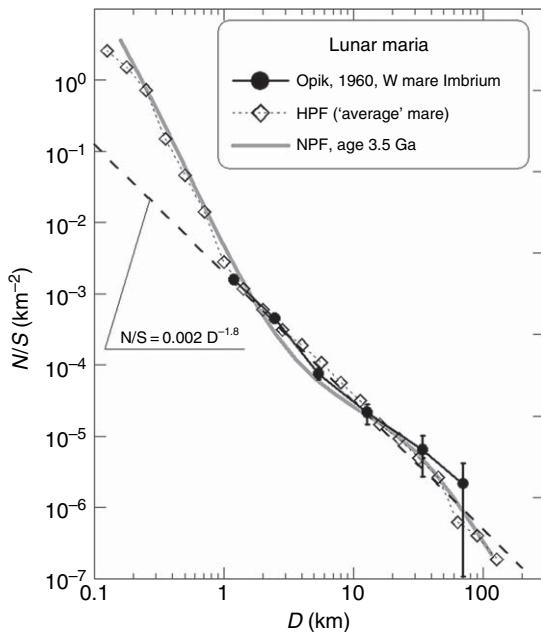


Figure 6 The early crater count (Öpik, 1960) in the lunar nearside in comparison with more extensive data published later.

telescopic images of the lunar nearside (Opik, 1960). Hartmann (1965) also reviewed telescopic crater counts and concluded that crater SFD matches asteroid SFD. The correspondence of these data with much more detailed study discussed below is amazing. Also, Öpik's early data set shows that the cumulative number of craters in the diameter range from ~ 1 to ~ 100 km is close to the $D^{-1.8}$ dependence.

The SFD of craters or projectiles is the relation between the object size (e.g., a diameter) and the number of such objects in a population under study. There are three main styles of the SFD presentation (Hartmann, 1965, 1969; Chapman and Haefner, 1967; Working_Group et al., 1979):

1. Cumulative form of SFD presents the number of craters larger than a given diameter, $N(>D)$.
2. Differential form of SFD presents the incremental number of craters in a given diameter bin (from the 'left' bin diameter to the 'right' bin diameter), $D_L < D < D_R$, as $\Delta N/(D_R - D_L)$.
3. Logarithmic form of the differential SFD presents the incremental number of craters in a given logarithmic D bin where $D_R/D_L = \text{constant}$, typically $\sqrt{2}$. The slope (in near-linear segments) equals that of the cumulative SFD.

4. Relative form of SFD presents the differential form, multiplied by D^3 :

$$R = D_{av}^3 (\Delta N/S)/(D_R - D_L) \quad [19]$$

where the geometrical mean interval diameter is defined as $D_{av} = (D_R \times D_L)^{0.5}$. The interval number of craters is divided by the area including counted craters making the R -value nondimensional. While one counts other kind of objects (e.g., asteroids) the R -value may be a dimensional one. The graphical representation of the $R(D_{av})$ relations is known as the ' R -plot'.

Cumulative SFD forms tend to smooth irregularities – deficiencies of craters in certain D bins. Differential forms, essentially histograms, have the advantage of showing the actual $N(D)$ behavior in each D bin, including turndowns caused by crater obliteration, and the log differential forms having the advantage of mimicking the cumulative $N(D)$ slope.

An ideal case to study the crater SFD is when a planetary surface is erased by some process and instantaneously begins to accumulate new craters. Before the crater degradation/obliteration processes change the population of these craters, the crater SFD reflects the production SFD or 'the standard distribution' of the projectiles. To understand this production function, many authors have tabulated and generalized a huge amount of data on lunar crater counts. Early results are summarized by Hartmann (1969, 1970a), Neukum et al. (1975), and Neukum (1983), and in a multi-author compilation of data by Hartmann et al. (1981). In this chapter, we concentrate on the lunar 'production' crater SFDs proposed by W. Hartmann and G. Neukum (see details in Neukum et al. (2001a)), and the update by Hartmann (2005).

10.06.4.1 Hartmann Production Function

To represent the crater SFD found on the terrestrial planets, Hartmann uses the log-incremental SFD representation with a standard $\sqrt{2}$ diameter bin size. Figure 4 in Hartmann (2005) is a proper illustration of the process of averaging of individual mare counts. We call his result the 'Hartmann production function', or HPF. The number of craters per square kilometers here is calculated for craters in the diameter bin $D_L < D < D_R$, where D_L and D_R are the left and right bin boundary and the standard bin width is $D_R/D_L = 2^{1/2}$.

The tabulated HPF is an assemblage of data selected by Hartmann to present the production function for one specific moment of time – the average time of lunar mare surface formation. For larger craters ($D > 4$ km), he used an exhaustive catalog of individual crater-by-crater diameters, measured by Arthur *et al.* (1963, 1965a, 1965b, 1966). Here the condition to have a fresh surface is satisfied by the fact that most lunar mare basalt samples have a narrow range of ages (e.g., 3.2–3.6 Gy; Stöffler and Ryder, 2001). Usage of the cratering chronology, discussed later in this chapter, helps to estimate the rate of accumulation of mare basalt flows (Hiesinger *et al.*, 2003). Using figure 13 from Hiesinger *et al.*, we plot the estimated cumulative volume of mare basalts, accumulated with age, in **Figure 7**. The median volume fraction (50%) of basalt flooding was reached ~3.4 Ga while the interval from 0.2 to 0.8 fraction occupied the model age interval from ~2.8 to ~3.5 Ga. This time interval may be used as an estimator of the ‘time slit width’ for Hartmann’s crater counts for an ‘average mare’ surface.

The tabulated HPF has been constructed by starting with a lunar curve that combines and averages crater counts in different mare areas. These lunar counts included data from the meticulous Arthur catalog which tabulated frontside lunar craters of $D > 4$ km. **Table 1** lists these Hartmann’s mare crater counts in each diameter bin together with Hartmann’s

scaled estimates of the Martian Neukum production function (NPF) (Hartmann, 2005). The total number of largest craters (**Table 1**) illustrates that during the last 3.5 Ga the Moon experienced ~10 impacts, creating craters 64–128 km in diameter, while ~100 craters larger than this diameter have been formed on the whole of Mars. The corresponding average distance between these craters is approximately 2000 km on the Moon and 1200 km on Mars. One can state that in the last 3.5 Ga the Moon and Mars have accumulated relatively sparse populations of largest craters.

Hartmann’s estimates for 1 Ga (the last column in **Table 1**) were scaled by him to any age younger than ~3.3 Ga assuming the constant cratering rate.

The HPF may be presented as a piece-wise three-segment power law (Ivanov, 2001; Hartmann, 2005):

$$\log N_H = -2.198 - 2.20 \log D_L, D_L > 64 \text{ km} \quad [20a]$$

$$\begin{aligned} \log N_H = -2.920 - 1.80 \log D_L, \\ 1.41 \text{ km} < D_L < 64 \text{ km} \end{aligned} \quad [20b]$$

$$\log N_H = -2.616 - 3.82 \log D_L, 0.3 < D_L < 1.41 \text{ km} \quad [20c]$$

Mare craters reach a saturation equilibrium areal density above $D \sim 300$ m and to define the production function for smaller craters one should use crater counts in areas, much younger than 3.5 Ga. Deriving Mars isochrones, Hartmann (2005) adopted a continuation of the curve by grafting on the shape at the small- D end of the NPF, discussed below. The NPF uses the –2.9 cumulative slope (Shoemaker *et al.*, 1970) giving lunar HPF-like SFD for small craters as:

$$\log N_H = -2.0 - 2.90 \log D_L, 0.01 < D < 0.125 \text{ km} \quad [20d]$$

Note that Hartmann (2005) includes effects of atmospheric shielding for small Martian impact craters.

10.06.4.2 Neukum Production Function

In the early time of lunar crater study there were a lot of discussions about the ‘best’ power law to present the impact crater SFD. The gradual accumulation of data resulted in the understanding that there is no unique power law, presenting the crater SFD in the whole range of crater diameters (Chapman and Haefner, 1967). Following the idea of a gradual SFD slope change with the crater diameter, Neukum in a series of publications – see summaries by Neukum (1983); Neukum and Ivanov (1994) – proposed an analytical function to describe the cumulative SFD of lunar impact craters. He showed that the production function had been more or less stable from Nectarian to

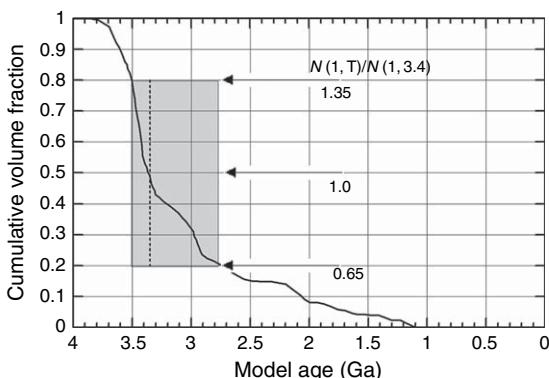


Figure 7 Cumulative volume accumulation of mare basalts estimated by crater count dating of individual basalt flows (Hiesinger *et al.*, 2003). The shaded rectangular outlines the ‘time slit’ for the accumulation of 60% of a total estimated volume. Median age of basalt accumulation is about 3.4 Ga, spanning from 3.5 (+30% of accumulation) to 2.8 Ga (~30% of accumulation). Estimated number of craters larger than 1 km is about $\pm 35\%$ of the $N(1 \text{ km})$ value for the median mare basalt age. This is a reasonably good estimator for the inherent accuracy of the HPF, combined from piecewise crater counts on various mare surfaces.

Table 1 HPF for the Moon and Mars

D_L	D_R	The Moon			Mars		
		Mare counts N_H per 1 km^2	Global number of large craters	N_H , 3.5 Ga	Global number of large craters	N_H , 3.5 Ga	N_H , 1 Ga
3.90 m	5.52 m			2.31 (4) ^b		1.24 (4) ^b	4.04 (3) ^b
5.52 m	7.81 m			1.33 (4) ^b		7.14 (3) ^b	2.33 (3) ^b
7.81 m	11.0 m			6.53 (3) ^b		3.50 (3) ^b	1.14 (3) ^b
11.0 m	15.6 m	$7.33 \pm 0.29 (1)^a$		2.62 (3) ^b		1.41 (3) ^b	4.58 (2) ^b
15.6 m	22.1 m	$2.42 \pm 0.12 (1)^a$		1.09 (3) ^b		5.85 (2) ^b	1.91 (2) ^b
22.1 m	31.2 m	$1.57 \pm 0.10 (1)^a$		3.81 (2)		2.04 (2)	6.66 (1) ^b
31.2 m	44.2 m	$5.82 \pm 0.48 (0)^a$		1.37 (2)		7.35 (1)	2.4 (1) ^b
44.2 m	62.5 m	$3.13 \pm 0.35 (0)^a$		5.40 (1)		2.90 (2)	9.44 (0) ^b
62.5 m	88.3 m	$1.87 \pm 0.27 (0)^a$		1.89 (1)		1.01 (1)	3.3 (0) ^b
88.3 m	125 m	$1.84 \pm 0.17 (0)^a$		6.95 (0)		3.73 (0)	1.22 (0) ^b
125 m	176 m	$1.06 \pm 0.13 (0)^a$		2.50 (0)		1.34 (0)	4.37 (-1) ^b
176 m	250 m	$8.08 \pm 0.55 (-1)^a$		8.40 (0)		4.51 (-)	1.47 (-1) ^b
250 m	353 m	$5.79 \pm 0.52 (-1)^a$		2.69 (-1)		1.44 (-1)	4.70 (-2) ^b
353 m	500 m	$1.04 \pm 0.06 (-1)$		7.91 (-2)		4.24(-2)	1.38(-2)
500 m	707 m	$3.23 \pm 0.14 (-2)$		2.30 (-2)		1.23(-2)	4.02(-3)
707 m	1 km	$1.14 \pm 0.05 (-2)$		6.60 (-3)		3.54(-3)	1.15(-3)
1 km	1.41 km	$1.64 \pm 0.09 (-3)$		1.76 (-3)		9.44(-4)	3.08(-4)
1.41 km	2 km	$5.74 \pm 0.46 (-4)$		7.32 (-4)		3.93(-4)	1.28(-4)
2 km	2.83 km	$2.82 \pm 0.25 (-4)$		3.92 (-4)		2.10(-4)	6.85(-5)
2.83 km	4 km	$1.21 \pm 0.11 (-4)$		2.10 (-4)		1.13(-4)	3.67(-5)
4 km	5.66 km	$8.31 \pm 0.37 (-5)$		1.13 (-4)		6.06(-5)	1.98(-5)
5.66 km	8 km	$5.10 \pm 0.29 (-5)$		6.05 (-5)		3.25(-5)	1.06(-5)
8 km	11.3 km	$2.56 \pm 0.21 (-5)$		3.24 (-5)		1.74(-5)	5.68(-6)
11.3 km	16 km	$1.67 \pm 0.18 (-4)$		1.74 (-5)		9.33(-6)	3.04(-6)
16 km	22.6 km	$5.51 \pm 1.06 (-6)$		9.29 (-6)		4.98(-6)	1.62(-6)
22.6 km	32 km	$4.29 \pm 0.94 (-6)$		4.98 (-6)		2.67 (-6)	8.71 (-7)
32 km	45.3 km	$2.24 \pm 0.68 (-6)$		2.67 (-6)		1.43 (-6)	4.67 (-7)
45.3 km	64 km	$2.04 \pm 0.72 (-6)$	~77	1.37 (-6)	~200	7.35 (-7)	2.4 (-7)
64 km	90.5 km	$2.14 \pm 2.14 (-7)$	~8	6.38 (-7)	~92	3.42 (-7)	1.12 (-7)
90.5 km	128 km	$2.14 \pm 2.14 (-7)$	~8	2.98 (-7)	~43	1.60 (-7)	5.21 (-8)
128 km	181 km	$1.00 (-7)^c$	~4	1.39 (-7)	~20	7.46 (-8)	2.43 (-8)
181 km	256 km	$4.69 (-8)^c$	~2	6.48 (-8)	~9	3.48 (-8)	1.13 (-8)
256 km	362 km	$2.18 (-8)^c$	~1	3.02 (-8)	~4	1.62 (-8)	5.28 (-9)
362 km	512 km			1.41 (-8)	~2	7.56 (-9)	2.47 (-9)
512 km	723 km			6.58 (-9)	~1	3.53 (-9)	1.15 (-9)
723 km	1024 km			3.07 (-9)	<1	1.65 (-9)	5.37 (-10)

^aDensity of craters is in the equilibrium state below the production function.^bThese values are for the production function estimates.

Copernican epochs (i.e., practically from more than 4 Ga until now). Neukum computed a polynomial fit to the cumulative number of craters, N , per square kilometers with diameters larger than a given value D . For the time period of 1 Gy, $N(D)$ may be expressed (Neukum, 1983) as

$$\log_{10}(N) = a_0 + \sum_{n=1}^{11} a_n [\log_{10}(D)]^n \quad [21]$$

where D is in kilometers, N is the number of craters with diameters $> D$ per square kilometers per billion

years ago, and the coefficients a_n are given in Table 2. Equation [21] is valid for D from 0.01 to 300 km.

Recently, the NPF was slightly reworked in the largest-crater part by careful re-measuring in the size range (Neukum *et al.*, 2001a). The time dependence of the a_0 -coefficient is discussed in the following subsection.

Although the HPF and NPF have some differences, both assume that the general shape of the SFD striking the Moon over the last 4 Gy was the same. A different point of view is given by R. Strom (Strom

Table 2 Coefficients in eqn [21]

a_i	'Old' N(D) ^a	'New' N(D) ^b	'New' N(D) Sensitivity ^c (%)
a_0	-3.0768	-3.0876	
a_1	-3.6269	-3.557528	± 3.8
a_2	+0.4366	+0.781027	± 3.9
a_3	+0.7935	+1.021521	± 2.5
a_4	+0.0865	-0.156012	± 1.6
a_5	-0.2649	-0.444058	± 0.88
a_6	-0.0664	+0.019977	± 1.3
a_7	+0.0379	+0.086850	± 0.78
a_8	+0.0106	-0.005874	± 1.8
a_9	-0.0022	-0.006809	± 1.8
a_{10}	$-5.18 \cdot 10^{-4}$	$+8.25 \cdot 10^{-4}$	± 5.6
a_{11}	$+3.97 \cdot 10^{-5}$	$+5.54 \cdot 10^{-5}$	± 24.1

^aNeukum, 1983.^bNeukum *et al.*, 2001a.^c'Sensitivity' is the coefficient variation which changes the N(D) value a factor of 2 up and down.

and Neukum, 1988; Strom *et al.*, 1992), who claims that the 'modern' (postmare) production function is quite different from that produced during the epoch of the late heavy bombardment. A more extensive treatment of this subject is published by Strom and Neukum (1988).

10.06.5 Impact Rate Estimates

Impact rate estimates for the earliest solar system planetary bodies (>3.8 Ga) may be obtained only from models. However, both Hartmann (1970a, 1970b) and Neukum (1983) have shown that crater densities at lunar landing sites of different ages require that the cratering rate increased dramatically (factor of 100) as we go back in time from ~ 3.2 to ~ 3.8 Ga.

Modern impact rates on planets are defined by orbital evolution of small bodies under weak gravity and nongravity forces, including resonances with large planets and effects of solar irradiation. The most violent cause of changes in the asteroid's physical state, orbital parameters, and rotations is collision with other small bodies. The extreme case of a collision is the so-called catastrophic collision, resulting in the asteroid disruption. By common definition, a catastrophic collision is an event, which produced the largest fragment with a mass of one-half of the initial asteroid mass (Greenberg and Hartmann, 1977; Greenberg *et al.*, 1978). For small asteroids (with diameters smaller than ~ 100 –500 m) the outcome of collision is controlled mostly by strength of the asteroid material. The effective strength is usually expressed in units of the specific

energy of the impactor per unit mass of the target body. For laboratory scale model rocky bodies the catastrophic specific energy is in the range of 300–2000 J kg⁻¹ (O'Brien and Greenberg, 2005). As the target size grows, most researchers assume a slow decline of this number due to the strength-scale effect – larger bodies in nature commonly contain more intrinsic defects (fractures, pores, inhomogeneities) decreasing the resistance to deformations and fragmentation (Love and Ahrens, 1996; Ryan and Melosh, 1998). For large asteroids (with diameters of ~ 500 m and larger) the self-gravity begins to play a role, keeping fragments of an impact together; the largest fragment may consist of numerous smaller fragments. The gravity self-keeping of fragments results in the concept of the presence of rubble piles – thoroughly shattered but not dispersed asteroids. The minimum of collision strength in various models is assumed for asteroids in the size range of 200–800 m (O'Brien and Greenberg, 2005).

In previous decades, collisions of asteroids were believed to be the main mechanism of their orbit evolution. In this concept, collision fragmentation of asteroids (1) produce small body population down to particles of dust size (which may be removed from the solar system by radiative pressure of the sunlight), and (2) change orbital parameters of fragments so that a fraction of fragments reaches planetary crossing (PC) orbits, producing meteoroids (found on Earth as meteorites) and crater-forming projectiles. The lifetime of these bodies at planet crossing orbits was assumed to be controlled by close encounters with planets (Arnold, 1965; Wetherill, 1967). From the modern point of view, this model gives too large estimates of the lifetime at

PC orbits (Dones *et al.*, 1999). Gravitational resonances are now seen to be a major factor in the planetary crosser delivery (as discussed below).

The frequency of giant impacts in the MB constrains the number of large crater-forming projectiles. The rarity of impacts that produce fragments larger than 1 km and divert them toward planets, as correctly pointed by Shoemaker, resulted in an early assumption that most of impact craters with diameters larger than ~ 30 km on terrestrial planets are formed by comets (Shoemaker and Wolfe, 1982; Shoemaker *et al.*, 1990). The recognition of a prevailing role of gravity forces (resonances) in repopulation of PC orbits from the MB (Farinella *et al.*, 1994) has destroyed Shoemaker's main argument for comet dominance. Modern estimate of the comet fraction in the impact cratering is about $10 \pm 5\%$ (Bottke *et al.*, 2002, 2005; Weissman *et al.*, 2002).

10.06.5.1 Gravitational Mechanisms

The critical mechanism in changing small body orbits is the resonant action of giant planets. There are two broad kinds of resonances (Morbidelli, 2002): (1) mean motion resonances (the periods of rotation around the Sun of two bodies give the ratio of integer numbers) and (2) secular resonances (the precession frequency of the asteroid's longitude of perihelion is equal to the integer number of secular frequency of the planetary system). The most powerful resonances with Jupiter creates gaps ('Kirkwood gaps') in the MB asteroid orbit distribution as a function of semimajor axis a (mean motion resonances with Jupiter are 3:1 ($a = 2.5$ AU), 5:2 ($a = 2.8$ AU), and 2:1 ($a = 3.2$ AU)). For the deflection of MB asteroids into PC orbits, two resonances are found to be the most important: 3:1 mean motion resonance with Jupiter and secular v_6 resonance with Saturn. These two resonances are studied in detail (Michel *et al.*, 2005). Many other resonances influence asteroid orbit evolution (Milani and Farinella, 1995; Morbidelli and Gladman, 1998). The action of resonances results in the quick increase or chaotic change of an asteroid's eccentricity, making its planetary crossers.

The Kozai resonance (which occurs when the precession of the asteroid's longitude of perihelion equals the precession of its node and keeps the quantity $a(1 - e^2)^{0.5} \cos i = \text{const}$, during a planet encounter) provides a sort of protection from close encounters. Bodies with low inclination and a semimajor axis close to that of Earth or Venus may enter the Kozai resonance, which causes the perihelion to librate around 0° or 180° . Moreover, the semimajor axis remains

almost constant, which protects the body from a collision with the planet (Milani *et al.*, 1989; Gladman *et al.*, 1995; Michel *et al.*, 1996; Michel and Thomas, 1996), at least temporarily.

10.06.5.2 Nongravity Mechanisms

The so-called Yarkovsky effect (Rubincam, 1998; Morbidelli and Vokrouhlický, 2003; Bottke *et al.*, 2006b) assumes that the solar heat, accumulated by the day side of an asteroid, re-radiates from the surface when a given area rotates to the night side. The effect depends on thermophysics of the thin surface layer, distance from the Sun, and a rotation of the asteroid. For a typical 1 km body the Yarkovsky effect results in the major semimajor orbit radius drift of the order of 10^{-4} AU per 10^6 years. The comparison of modeled Yarkovsky effect with astronomical observation is still limited to a few asteroid observations (Chesley *et al.*, 2003) and statistical investigations of asteroid family evolution (Vokrouhlický *et al.*, 2006a, 2006b). Yarkovsky effect operates slowly, so it is important for the MB evolution and delivery of small bodies to resonant orbits.

The so-called Yarkovsky–O'Keefe–Radzievskii–Paddack (YORP) effect (Rubincam, 2000; Cadek and Vokrouhlický, 2004) explains how the solar irradiation may change small body spin state (spin up or spin down) by the reflection and re-radiation of absorbed sunlight.

Once an object becomes PC, the main disturbing events are encounters with terrestrial planets, and the Yarkovsky effect has no time to build up the appreciable input for the orbit evolution (Michel and Yoshikawa, 2005).

10.06.5.3 Impact Probability

The impact probability of NEOs is a parameter that connects the observed NEO population and the number of actual impact events on the target planet, measured as the recent cratering rate (see Section 10.06.7), or as the rate of bolides, recorded as light flashes (meteors or bolides fireballs) in the terrestrial atmosphere. If the whole population of NEO in a steady state is N_{NEO} , and the average (for the whole NEO population) probability of collision with the target planet is P_{col} (measured in yr^{-1} per one object), the impact rate, R_{col} , defined as number of impacts (collisions) in the specific time period (e.g., per year) is

$$R_{\text{col}}(\text{yr}^{-1}) = N_{\text{NEO}} \times P_{\text{col}} \quad [22]$$

In a standard approach the value of P_{col} is estimated with Öpik-like formulas (Wetherill, 1967; Öpik, 1976; Bottke and Greenberg, 1993) provided that the main orbital parameters for NEO and the target planet, a, e, i are constant in time and all angles defining mutual position of orbital planes are equally probable. Typical values of the Öpik-like P_{col} are in the range of $\sim 10^{-11}\text{--}10^{-8} \text{ yr}^{-1}$ and during a typical lifetime of the individual NEO of 10^7 yr^{-1} the probability of impact is very low. Moreover, each NEO changes its orbit in a very wide zone of the phase space during its lifetime. Hence, eqn [22] should be treated as an estimate for the specifically populated orbit set provided the population is in a steady state: (1) if an Apollo object, for example, changes its orbit to the Amor-type orbit, another NEO changes its orbit from Amor-type to Apollo type (see Milani *et al.* (1989)); (2) the number of objects leaving the NEO family (striking Sun and planets or being ejected out of the inner solar system) are replenished with new NEOs arriving from the MB orbits.

Most of the astronomically observed NEOs are tested by numerical orbit integration to be estimated as a potentially hazardous object (PHO). The list of PHO close encounters with Earth is published by several groups. None of the tested NEOs, according to the numerical orbital motion integration, show a chance of striking the Earth or Moon in the historically short time. The forward integration (Morrison *et al.*, 2002; Harris, 2003; Stuart and Binzel, 2004) of Earth crossing NEO with $H < 18$ produces the statistics of close approaches (distance $< 0.1 \text{ AU}$) and converts it into the average impact probability of all NEOs (bodies with $q < 1.3 \text{ AU}$) as low as $1.7 \times 10^{-9} \text{ yr}^{-1}$ per one NEO, which is equivalent approximately to $3.4 \times 10^{-9} \text{ yr}^{-1}$ per one Earth crossing NEO.

The estimate of the impact (collision) probability for a given asteroid is the nontrivial problem, as small PC bodies are in chaotic motion: small differences in initial orbit parameters result in a set of very different evolution of orbits in a long timescale. A recent example is the evolution of the asteroid Itokawa (Michel and Yoshikawa, 2005). The direct usage of Öpik–Wetherill probability formulas gives impact probability of 0.8×10^{-9} and the impact velocity 12.5 km s^{-1} , while numerical integration of 39 ‘clones’ (each clone differs by the last digit in a, e, i parameters) result in four impacts during simulated 100 Ma (Michel and Yoshikawa, 2005) giving an estimate of collision $P = 4$ impacts per 39 clones per 10^8 years $\approx 1 \times 10^{-9} \text{ yr}^{-1}$. The chaotic evolution of the

orbits of a given small body demands the huge volume of integrated variants before some reliable average estimates could be formulated.

However, in parallel with upcoming numerical simulations, Öpik-type estimates for a large population of asteroids, based on cratering rate derivations, may be treated as the average probability of steadily populated PC orbits. The problem here is that the subset of orbits having the largest collision probability simultaneously are orbits with shortest lifetime: the high probability of a close encounter gives large chances both for collisions and for an abrupt orbit change during the close encounter. Hence, there is only a small chance to observe an NEO in these orbits. Partial solution of this problem may be found in the analysis of real collisions of small (hence frequent enough) bodies, burned in the atmosphere and observed as bolides (Morbidelli and Gladman, 1998). Systematic photographic surveys of bolides have been made in different countries in different periods of time (McCrosky *et al.*, 1978; Ceplecha, 1987; Halliday, *et al.*, 1996; Oberst *et al.*, 1998; Borovicka *et al.*, 2005). Registration of each event from several observational points allows researchers to estimate orbital parameters just before the atmospheric entry. Using these orbital parameters, one can make a formal calculation of the Öpik-type collision probabilities and compare these estimates with similar values for astronomically observed NEOs. In principle, these data may be different, as only NEOs with diameter of $\sim 1 \text{ km}$ and larger have close to complete statistics, while statistics of bolides are for small bodies with the effective diameter of a few meters and smaller. In addition, there are many inherent observational biases for bolide observations as a statistical database (Halliday *et al.*, 1996). To reduce severity of possible biases we make estimates for bolides with various minimum photometric mass (0.1, 1.0, and 10 kg) from the database for bolides from many surveys (Lindblad *et al.*, 2003).

To check the idea that small number of meteoroid collision with Earth occurs from a small number of orbits with high impact probability, data for NEOs and bolides are sorted in the order of increasing P_{col} with calculation of the accumulated average $\langle P_{\text{col}} \rangle$ for the meteoroid ensemble:

$$\langle P_{\text{col}}(n) \rangle = \frac{\sum_{i=1}^n P_{\text{col},i}}{n} \quad [23]$$

where i is the number of the orbit ($i = 1, N$), and $P_{\text{col},i} < P_{\text{col},i+1}$. In this construction first most-low-probability orbits are summed and the average value

grows as $P_{\text{col},i}$ monotonically increases with the number i in the sorted list. If a low number of orbits with higher collision probability give disproportional large input, the function $\langle P_{\text{col}}(n) \rangle$ would rise when n approaches the total number of recorded events, N . **Figure 8** presents the relation between $\langle P_{\text{col}}(n) \rangle$ and $P_{\text{col},m}$, the largest probability in the partial sum in eqn [23] for Earth crosser objects in asteroid-like orbits ($T_J > 3$) with absolute magnitudes $H < 15.5$ (close to completeness of the astronomical survey), $H < 18$ (roughly >1 km in diameter) objects, and for fireballs with photometric masses more than 0.1 and 10 kg. The

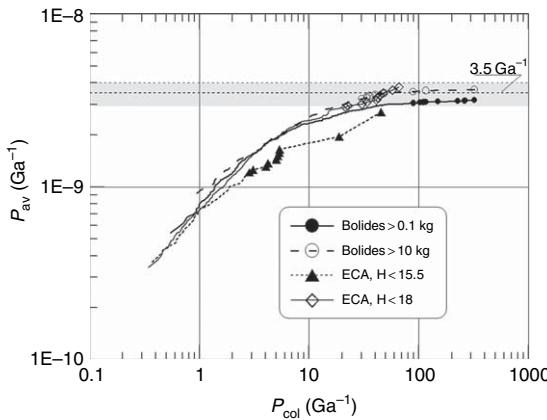


Figure 8 The accumulated sliding average for Earth-crossing asteroids (ECAs) and bolides, constructed by the ascending sorting of orbits on the collision probability (eqn [23]). All curves begin to be flat for rarely populated orbits with large collision probabilities. The average probability for the whole observed population is about $3.5 \pm 0.5 \text{ Ga}^{-1}$ per one Earth crosser.

absolute magnitude of asteroids, H , is defined as the apparent magnitude that the object would have if it were one astronomical unit (AU) from both the Sun and the Earth and at a phase angle of zero degrees. One can see that the function $\langle P_{\text{col}}(n) \rangle$ for $H < 15.5$ bodies and fireball meteoroids grows smoothly to the average per-object collision probability of $3.5 \pm 0.5 \text{ Ga}^{-1}$. These curves go flatten at largest probabilities, converging to the limiting average probability. For each data set 10 events with the largest collision probability are marked, to illustrate their importance. Data for $H < 18$ objects may be corrected for the incompleteness of astronomical survey for these bodies. If one assumes that still unrecognized $H < 18$ bodies are on orbits with small collision probabilities (low probability of close encounters), the sum in eqn [23] should be divided by larger number of bodies as assumed low collision probabilities do not make an essential input into the sum of individual orbital collision probabilities.

Celestial mechanics reveals two broad classes of orbits for small bodies in the solar system depending on whether the orbit allows close encounters with Jupiter or not. The astronomically recent close encounters result in the specific value of Tisserand parameter,

$$T_J = a_J/a + 2 \cos i [(1-e^2)(a/a_J)]^{1/2} \quad [24]$$

where a , e , i are correspondingly semimajor axis, eccentricity, and inclination of the small body orbit, and a_J is the semimajor axis of Jupiter.

Table 3 summarizes Öpik-type estimates for real collision orbits compared with similar estimates for large Earth crossers with $H < 18$. Estimates show that for objects (meteoroids) in asteroid-like orbits ($T_J > 3$)

Table 3 Estimates of collision probability and average impact (pre-atmospheric) velocity of bolides from meteor ground survey (IAU MDC Photographic Database, Version 2003)

	Known Earth-crossers		Observed bolides		
	$H < 18$	$m_{\text{ph}} > 10 \text{ kg}$	$m_{\text{ph}} > 1 \text{ kg}$	$m_{\text{ph}} > 0.1 \text{ kg}$	
$T_J > 3$					
N	401	74	152	227	
$P_{\text{col}}, \text{yr}^{-1}$	3.8×10^{-9}	3.7×10^{-9}	3.2×10^{-9}	3.2×10^{-9}	
$\langle U \rangle$	19.0	19.0	19.8	20.3	
$2 < T_J < 3$					
$N(\% \text{ of total})$	54(12%)	18(20%)	48(24%)	106(32%)	
$P_{\text{col}}, \text{yr}^{-1}$	0.80×10^{-9}	0.88×10^{-9}	0.90×10^{-9}	0.93×10^{-9}	
$\langle U \rangle$	26.4	27.5	25.2	25.5	
All objects with $T_J > 2$					
N	455	92	200	333	
$P_{\text{col}}, \text{yr}^{-1}$	3.4×10^{-9}	2.3×10^{-9}	2.0×10^{-9}	1.8×10^{-9}	
$\langle U \rangle$	19.2	20.7	21.6	22.1	

the average collision probability and average collision velocity for meteoroids, created bolides entering the terrestrial atmosphere, are close to values for observed Earth crossers: $P_{\text{col}} \sim (3-4) \times 10^{-9} \text{ yr}^{-1}$ per body and $\langle U \rangle \sim 19-20 \text{ km s}^{-1}$ which corresponds well with results of orbital integrations and bias corrected estimates (Morrison *et al.*, 2002; Stuart and Binzel, 2004). Objects on comet-like orbits ($2 < T_{\text{J}} < 3$) have a tendency to give larger percentage of observed bodies with the size (mass) decreasing from $\sim 10\%$ for NEOs with $H < 18$ to $\sim 30\%$ for bolides with small photometric masses. The observational bias cannot be excluded, but the tendency deserves a further investigation. The average collision probability and impact velocity for objects on comet-like orbits are: $P_{\text{col}} \sim (0.8-0.9) \times 10^{-9} \text{ yr}^{-1}$ per body and $\langle U \rangle \sim 25-26 \text{ km s}^{-1}$. The average values for all Earth-crossing asteroids and largest bolides are slightly different due to apparent difference in percentage of bodies, striking Earth from asteroid-like and comet-like orbits.

We can conclude that collision probability estimates based on fireball and observed asteroid databases are consistent, and that fireball data, resulted from real impacts, gives an additional support to celestial mechanics results.

Similar estimates based on observed asteroids give approximate estimates of collision probabilities for other terrestrial planets (Hartmann, 1977; Hartmann *et al.*, 1981; Ivanov, 2001; Neukum *et al.*, 2001b). **Table 4** gives an example of collision probabilities and average impact velocities estimates, based on known population of planetary crossers with $H < 17$.

One can states that the ‘bolide ratio’ within a factor of 2 is the same for Mercury, Venus, Earth, and the Moon. Mars, being close to the main asteroid belt, suffers 3–4 more frequent impacts. However, the larger surface gravity and smaller average impact velocity on Mars in comparison with the Moon result in smaller crater sizes, created by projectiles of a given size. Hence, the ratio of ‘cratering rate’, the frequency of the same size crater formation at the unit surface area on Mars and the Moon is smaller than bolide ratio. Thus, formation of an impact crater of a given size on Mars needs larger, consequently less numerous, projectiles than on the Moon (Ivanov, 2001; Hartmann, 2005).

10.06.6 Extraterrestrial Surface Dating

10.06.6.1 Principles of Surface Dating by Crater Counting

Principles of extraterrestrial surface dating have been formulated decades ago (Hartmann, 1977; Hartmann *et al.*, 1981) as a sequence of steps:

1. Count readily recognizable impact craters on units of interest.
2. Use a baseline crater production rate as measured in the Earth–Moon system primarily from craters on lunar surfaces dated from Apollo and Luna sample return missions (Stöffler and Ryder, 2001).
3. Estimate relative impact rate on the planet of interest primarily from observed interplanetary

Table 4 Average impact probability and average impact velocity on terrestrial planets for observed planetary crossers $H < 17$, listed in the MPC catalog (file ‘astorb.dat’) in November 2005

Planetary body	$N_{H < 17}$	Average impact rate ^a ($10^{-15} \text{ km}^{-2} \text{ yr}^{-1}$)	Average collision probability per one body (planetary crosser) (Ga ⁻¹)	Average impact velocity (km s ⁻¹)	R_b ^b
Mercury	34	0.715	1.574	35.4	0.92
Venus	89	0.876	4.53	24.2	1.12
Earth	194	1.31	3.443	19.3	1.68
The Moon	194	0.781	0.157	17.5	1
Mars, current orbit, $e = 0.0934$	2680	3.85	0.208	9.4	4.93 ^d
Mars, $e = 0.05^c$	1433	2.42	0.244	10.2	3.10 ^d
Mars, $e = 0.01$	869	2.0	0.336	10.5	2.58 ^d

^aAverage impact rate is the global number of impacts per a time unit, divided by the surface area of a planet.

^b‘Bolide ratio’ is the ratio of the impact rate of projectile of the same size per unit area to the same value for the Moon (for the Moon $R_b = 1$ by definition).

^cThis value of eccentricity is close to the average value for a time span of 10 Ma and longer still limited to the available Mars orbit modeling duration (Ward, 1992).

^dNot corrected for the incomplete observation of Mars crossers (see Ivanov, (2001) for an example of the correction).

populations and dynamical models of asteroid scattering (Hartmann, 1977; Marchi *et al.*, 2005).

4. Use the impact scaling law to convert the assumed impact rate into crater formation rate taking into account the average impact velocity, surface gravity, and target planet material properties (presence of volatiles, strength, porosity etc.).

5. Estimate the surface age (the crater population accumulation period) dividing the number of counted craters in each crater diameter bin by the assumed crater formation rate.

Each of these steps add an uncertainty to the final number which currently may be estimated as a factor of 2 at best and as an order of magnitude in worse cases. Nevertheless, step (1) only (crater count) often gives a very useful understanding of relative ages of different geological units of the same planet.

When dating surfaces with the age of ~ 3.3 Ga and older, one more factor should be taken into account. As it was demonstrated for the Moon, even before Apollo missions (Hartmann, 1966b), the late bombardment era is characterized with much larger cratering rate. Hence, using the lunar cratering chronology we implicitly assume that the decaying early bombardment flux was the same for all terrestrial planets. The validity of this assumption partly

depends on the assumed nature of the heavy bombardment period (Levison *et al.*, 2001; Hartmann *et al.*, 2002; Gomes *et al.*, 2005). Pending better understanding, the assumption of the similar decay rate from ages ~ 4 to ~ 3.3 Ga and approximately constant bombardment flux looks like a reasonable first guess. The illustration of the general behavior of the bombardment flux is presented in Figure 9 for the Moon, comparing ages of returned samples and crater counts in the vicinity of landing sites.

The constancy of the bombardment flux is an approximation in two respects. First, catastrophic collision of large MB asteroids may have temporarily increased the bombardment flux into planetary crossing orbits. Second, there is a hypothesis (based on limited observational data) about gradual decay of the bombardment flux during the last 3 Ga (Hartmann *et al.*, 2007; Quantin *et al.*, 2007). More work in the celestial mechanics is needed to add a plausible explanation to this hypothesis from the point of view of minor body orbital and collision evolution.

In addition to calibration by using dated lunar provinces, the assumption of a common bombardment history for all terrestrial planets may also be used as a foundation for the universal comparative crater chronology for all terrestrial planets as was proposed by Neukum and Hiller (1981). This approach assumes

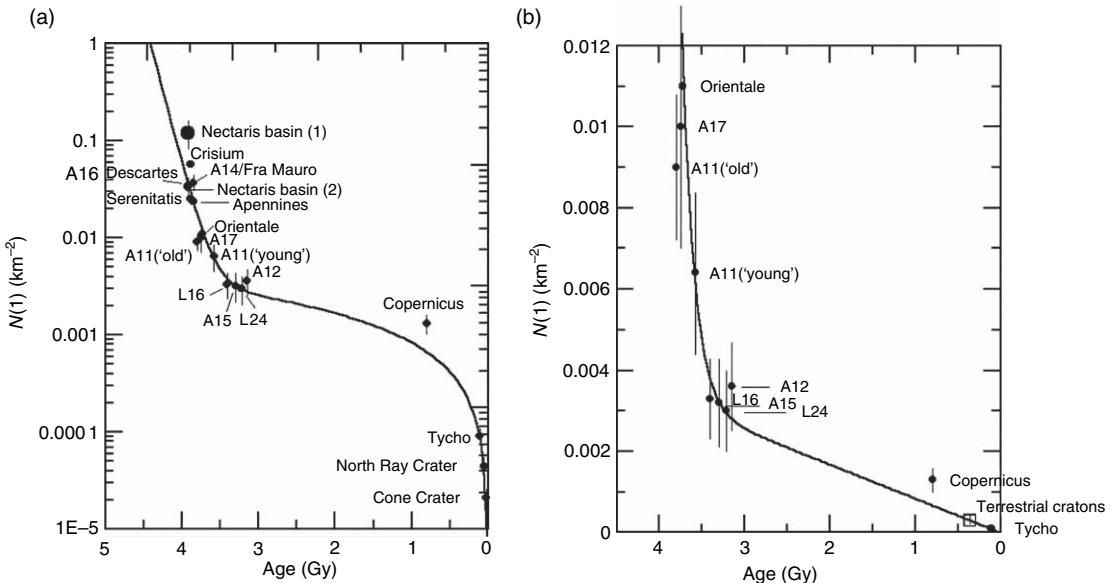


Figure 9 The accumulation rate of lunar impact craters, described as number of craters with $D = 1$ km, $N(1)$, estimated by fitting real crater counts with a production function, (a) data are in a semi-logarithmic plot, (b) data are plotted in a linear scale for younger crater population. Crater counts and sample ages are from Neukum (1983), Neukum and Ivanov (1994), and Stöffler and Ryder (2001). Reprinted from Neukum G, Ivanov BA, and Hartmann WK (2001a) Cratering records in the inner solar system in relation to the Lunar reference system. In: Kallenbach R, Griss J, and Hartmann WK (eds.) *Chronology and Evolution of Mars*, pp. 55–86. Dordrecht: Kluwer Academic Press, with permission from Springer.

that all giant multi-ring basins were formed in the same epoch during the late bombardment. Dating of these basins with counting of postbasin superimposed craters one can ‘tune’ the lunar time dependence of the bombardment rate to the planet of interest. Absolute ages may be derived making use of the idea of the ‘marker horizon’, that is, formation of the youngest basins, such as Gilgamesh on Ganymede, about 3.8 Ga (Neukum *et al.*, 1998). In general, this approach gives results similar to estimates based on asteroid bombardment rate comparison – see examples for Mars and Mercury (Hartmann and Neukum, 2001; Neukum *et al.*, 2001b). However, this technique does not apply for Venus where cratering records of the late bombardment period have been erased by intensive endogenic processes.

This was recognized as early as the mid-1960s by Öpik who spoke of effects of Martian erosion, and by Hartmann (1966a), who developed the concept of diameter-dependent ‘crater retention ages’, to emphasize that crater count ages date the survival time of craters in each diameter bin. The concept is analogous to gas retention ages of rocks. In ideal conditions, such as a well-preserved lava flow on an airless planet, the crater retention age may equal the formation age of the rock unit. But in erosive or depositional regimes, it may measure only the survival time against obliteration, as a function of size, and thus give information on rates of geological resurfacing processes. Similarly, a population of fresh, sharp craters on an exhumed surface may characterize the time since exhumation.

The most universal (however, very model-dependent) approach to calibrate the crater chronology model for a planet is to compare crater formation rate with the lunar one. Fairly simplified calibrations were made by Hartmann (1977) and Hartmann *et al.* (2001). The more rigorous approach (Ivanov, 2001; Ivanov *et al.*, 2001) includes treatment of:

1. lunar impact velocity distribution;
2. projectile population SFD from the lunar cratering records;
3. impact probability and impact velocity distribution in a target planet; and
4. calculation of a crater SFD on the target planet using the lunar-derived projectile SFD.

Step (1) may be done with probability estimates for NEA orbits resulting in a function f_V to define the probability of impact with a given velocity (NEA is near-Earth asteroids on orbits which cross the Earth

orbit.). Step (2) in a full treatment demands solution of the integral equation

$$\frac{dN}{dD} = \int_{U_{\min}}^{U_{\max}} dU \int_{\phi_{\min}}^{90^\circ} d\phi \frac{dN}{dD_p} \frac{dD_p}{dD} f_U(U) f_\phi(\phi) \quad [25]$$

to find the differential SFD for projectiles, dN/dD_p , from the measured SFD for craters, dN/dD , provided we know the probability of impacts, f_U , in the impact velocity range from U_{\min} to U_{\max} , and the cratering scaling law, the dependence of the crater diameter D on the projectile size, D_p , impact velocity, U , and impact angle ϕ : $D = D(D_p, U, \phi)$. The frequency of oblique impacts, f_ϕ , is known from theoretical considerations (eqn [17]). The current knowledge of the scaling law is discussed in Section 10.06.3.

In a simplified approach, as in the early publications, one can compute the average impact velocity $\langle U \rangle$ (Table 4) and assume that all impacts occur at the most probable impact angle of 45° . In this case the number of projectiles, created craters of diameter D is the same as number of projectiles with diameter D_p , provided the scaling law $D = D(D_p, \langle U \rangle, 45^\circ)$ is known. Hence, $dN/dD_p = dN/dD$. In both the simplified and ‘full-treatment’ (eqn [25]) approaches we still assume that all projectiles have the same density. In reality the known range of projectile density is in the range from ~ 500 (porous comet nucleus) to $\sim 7000 \text{ kg m}^{-3}$ (iron meteoroids), including most practically important values $\sim 2700 \text{ kg m}^{-3}$ for S-type asteroids and $\sim 1300 \text{ kg m}^{-3}$ for dark (e.g., C-type) asteroids (Britt *et al.*, 2002; Stuart and Binzel, 2004). Recent Hayabusa mission to asteroid Itokawa reveals the density of this $\sim 500 \text{ m}$ S-type body as low as ~ 2100 to 2300 kg m^{-3} . (Fujiwara *et al.*, 2006).

Step (3) again needs to use model or observational information about orbits of small bodies crossing the orbit of a target planet. While the average impact velocity estimated by different groups for various terrestrial planets are close (see Table 4; the average impact rate (the number of impacts of bodies of the same size per unit time per unit area of the target planet surface)) requires knowledge of asteroid/comet orbital dynamics much better than current understanding. However, the comparison of pioneer publications (Hartmann, 1977; Shoemaker, 1977; Hartmann *et al.*, 1981) based on a dozen of small body orbits known at that time with recent estimates included hundreds of orbits does not demonstrate a dramatic difference in estimates. Starting with observed small body orbits one can make estimates

of the average impact rate ratio on a given planet relative to the Moon, illustrated in [Table 4](#).

Step (4) applies cratering scaling laws taking into account the target planet surface rock properties and gravity acceleration. The full-treatment uses eqn [25] to compute the crater SFD. In a simplified treatment the average impact velocity and impact angle of 45° may be used.

Below we illustrate available data on the relative cratering rate for terrestrial planets.

10.06.6.2 Earth–Moon System

Earth and the Moon are subjected approximately to the same bombardment flux, so the Earth/Moon comparison is a good starting point to discuss problems of interplanetary comparison of cratering rates. For the Moon we have reasonably reliable cratering counts, in a few cases ‘calibrated’ with radiometric ages of returned samples. The most extensive crater counts are available for lunar mare surfaces, while crater records for younger surfaces are typically sparse. For Earth we have relatively small number of craters and heavy erosive losses of small ones, also we do have good age estimates from radiometric dating of impact melt and from stratigraphy. In addition, the data on frequency of meteoroid’s entry to the terrestrial atmosphere give a snapshot of the modern bombardment rate.

The Earth–Moon bombardment history comparison is tightly connected to the general history of the Earth–Moon system formation about ∼4.5 Ga. It is now believed that the Moon has been formed near the Earth’s Roche limit at a distance of ∼3 Earth radii, and has been moved to the current orbit at about 60 Earth radii by tidal forces (Ruskol, 1966; Touma, 2000). The impact rate ratio in the Earth–Moon system partially depends on the Earth–Moon distance (Bandermann and Singer, 1973), defining the modest near-side/far-side and leading/trailing asymmetry of impact rates decreasing for larger approaching velocity. The severity of this asymmetry depends also on the approach velocities of celestial projectiles. Attempts to reveal a (small) lunar cratering asymmetry are continuing (Morota and Furumoto, 2003) but theoretical efforts recently have been to improve our understanding of the lunar cratering asymmetry issues (Le Feuvre and Wieczorek, 2005, 2006; Gallant and Gladman, 2006; Morota *et al.*, 2006) but do not give very consistent results. The most robust of these resent studies (Gallant and Gladman, 2006) favor a minor (a few

percent) enhancement of the cratering rate in the leading hemisphere of the Moon. At the current level of lunar surface dating, we believe such effects can be neglected.

The comparison of available data for the Earth–Moon system in general results in the conclusion, that the modern (10 years scale) and geologically recent (100–500 Ma) bombardment flux is the same as proposed for the 3 Ga averaged lunar cratering rate within a factor of 2 (Grieve, 1984; Grieve and Shoemaker, 1994; Grieve and Pesonen, 1995; Werner *et al.*, 2002; Stuart and Binzel, 2004). However, some details of this comparison point to the differences looking like a systematic misfit. Below we illustrate possible reasons for this and possible ways of the reconciliation.

[Figure 10](#) summarizes observational data about cumulative probability of collision of small bodies with a character diameter larger than D_p with the Earth (or entry to the terrestrial atmosphere). Small diameter data (D_p in the range of 10–20 cm) are based on fireball Earth-based registration (Halliday *et al.*, 1996). Data for larger bodies (D_p in the range of 1–10 m) are derived from satellite registration of bolide flashes (Brown *et al.*, 2002) where projectile masses are estimated from the conversion of light flash energy to a kinetic energy with the following conversion to sizes assuming the average impact velocity of 20 km s⁻¹ and average density of 2700 kg m⁻³. From the opposite end of a size range ($D_p \geq 5$ km) the estimate is derived from observed NEA population and the model of albedo distribution among NEAs (Binzel *et al.*, 2004; Stuart and Binzel, 2004). In the intermediate size range of 10 m < D_p < 5 km the incompleteness of astronomically observed NEAs increased toward the small sizes. Here one should use various models, converting the (small) number of observed bodies into the full population assuming different estimators of the completeness of observations (D’Abramo *et al.*, 2001; Stuart, 2001; Bottke *et al.*, 2002; Stuart and Binzel, 2004). All models demand a model for albedo distribution among NEAs to be converted into diameters, and an assumed average collision probability. The latter is the most unreliable procedure, as one needs to assume orbital parameters of still unobserved bodies. The correlation with terrestrial impact records uses scaling laws to convert the terrestrial crater size into a projectile mass for an assumed average impact velocity, and the average density of projectiles, formed terrestrial craters. To decrease the number of assumptions the data for terrestrial cratering records are replotted in [Figure 11](#) as the global

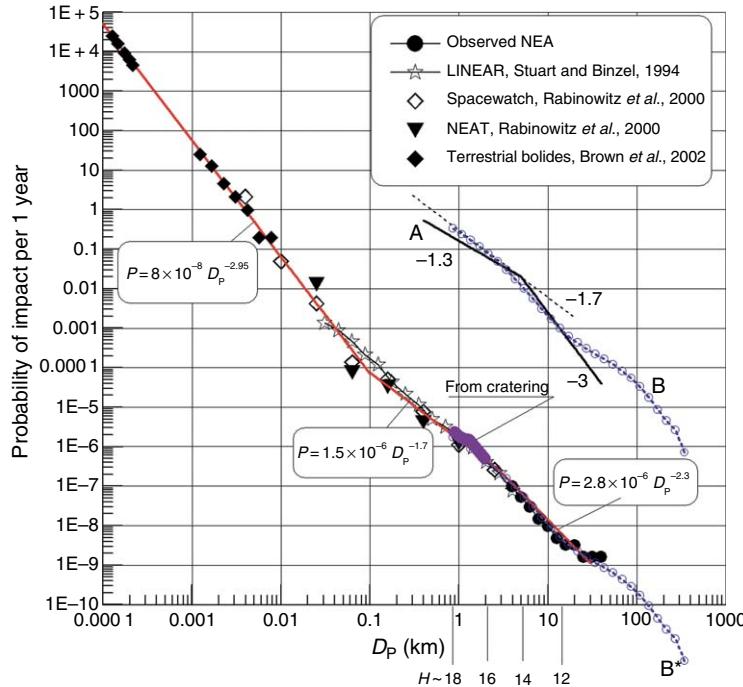


Figure 10 Probability estimates compendium for impacts over the total Earth's surface for small bodies larger than D_p . Data for bolides and fireballs ($D_p < 10$ m) are taken from Halliday *et al.* (1996) and Brown *et al.* (2002); observed large NEA are from 'astorb.dat' file (October 2005); estimates for small NEA are derived from model corrected to the observational completeness (Rabinowitz *et al.*, 2000; Stuart, 2001; Stuart and Binzel, 2004). Thick black segment 'From cratering' is from Hughes (2000) data on terrestrial cratering rate in the last 100 Ma converted to D_p for average impact velocity. MB SFD estimates (plotted here in the arbitrary vertical position) are from Ivezic *et al.* (2001) in the form of two-segment power law (A) with exponent index -1.3 and -1.7 proposed by Ivezic *et al.* and from Jedicke *et al.* (2002) (curve B). The latter is scaled to the NEA probability curve for comparison, showing the similarity of SFD of NEA and MB asteroids with limits of available data and models at $H < 18$.

probability of the formation of an impact crater with diameter larger than D . Here observations include the direct data on terrestrial crater ages and spacing (Grieve, 1984; Shoemaker and Shoemaker, 1990; Grieve and Shoemaker, 1994). NEA estimates convert kinetic energy frequency model (Stuart and Binzel, 2004) into terrestrial impact crater diameter with scaling laws discussed above (eqns [8] and [15]). Within a factor of 2 both astronomical (snapshot) and geologic (average for 100–400 Ma) estimates of the terrestrial cratering rate are similar, confirming the relative constancy of the bombardment flux to Earth in recent time (<0.5 Ga).

With a lot of precautions about the data accuracy, we can derive an approximate composite summary of the current bombardment rate as the global terrestrial probability of impacts of small bodies with an average density of $\sim 2700 \text{ kg m}^{-3}$ (bolide and fireball data) and a mix of larger objects (following the proportion of 'dark' objects of

low density and 'bright' objects with higher density proposed by Stuart and Binzel, 2004) in the form

$$P(>D_p) \approx 8 \times 10^{-8} D_p^{-2.95}; D_p < 100 \text{ m} \quad [26a]$$

$$P(>D_p) \approx 1.5 \times 10^{-6} D_p^{-1.7}; 100 \text{ m} < D_p < 1 \text{ km} \quad [26b]$$

$$P(>D_p) \approx 2.8 \times 10^{-6} D_p^{-2.3}; 1 \text{ km} < D_p < 20 \text{ km} \quad [26c]$$

Equation [26] presents an attempt to present the general nonpower law SFD for asteroids as a set of power law segments similarly to the lunar crater HPF [20]. The main guideline in the construction of eqn [26] is to pass the power law segments close (factor of 2) to all published data, and any extrapolation of these equations is completely invalid. Published data, shown in Figure 10, also offer the power-law fits. However, all main belt data may be well scaled to the NEA curve at least to $H < 18$ magnitudes (small bodies approximately of 1 km in diameter). The nongravity effects in modeled MB to NEA asteroid evolution predict the steepening of NEA SFD versus MB SFD due to larger small

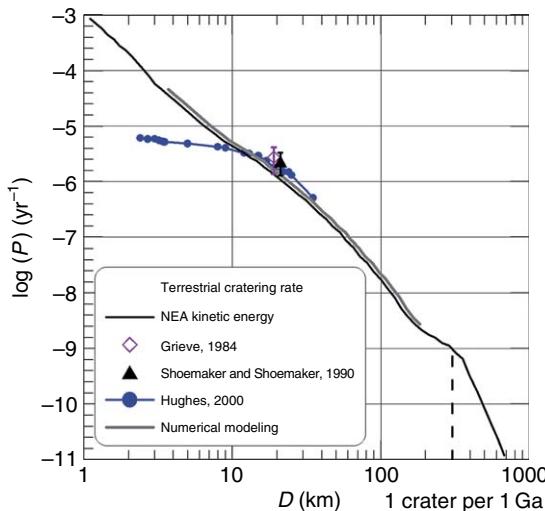


Figure 11 Cratering rate estimates for large terrestrial craters in comparison with Stuart and Binzel (2004) model of frequency distribution of kinetic energy of NEA impacts. The kinetic energy is converted into the crater diameter for the average impact velocity of 20 km s^{-1} using the scaling laws derived from experiments (eqns [8] and [15] – black curve) and with the scaling law derived from the numerical modeling of terrestrial craters (eqn [16] – gray curve). Within a factor of 2 the currently observed (snapshot) probability of the new crater formation is close to geologically observed cratering rate, averaged on the time period of 100–400 Ma (Grieve, 1984; Shoemaker and Shoemaker, 1990). To some extent the NEA kinetic energy estimates depend on the assumed albedo distribution among NEAs on various orbits.

asteroid ‘mobility’ in the phase space due to Yarkovsky and YORP effects (Morbidelli and Vokrouhlický, 2003). These effects are believed to dominate in decameter size range (e.g., figure 7 in Jedicke *et al.*, 2002), but affect different taxonomic types differently (Hartmann *et al.*, 1999). Nongravity effects on NEA SFD are still visible only as a tendency for $H < 18$ (Figure 10), while radiative models offer great progress in the understanding of asteroid family evolution (Bottke *et al.*, 2006a; Vokrouhlický *et al.*, 2006a, 2006b).

Here we apply the ‘composite’ bombardment flux (eqns [26a]–[26c]) only to make the direct recalculation of estimated terrestrial bombardment rate into lunar cratering rate estimates. Assuming the average impact velocities as 20.9 and 19.2 km s^{-1} for Earth and the Moon correspondingly, effective gravity-enhanced cross sections as 7.12×10^8 and $3.85 \times 10^7 \text{ km}^2$ (Stuart and Binzel, 2004), one can estimate the Earth/Moon bolide ratio as 1.42 (note the difference with $R_b = 1.68$ estimated from observed orbits of NEA with $H < 17$ – see Table 4). Using scaling laws discussed above we can calculate

the lunar production function and compare it with the observed one. Comparison of the modern Earth/Moon projectiles with lunar production functions (NPF and HPF) in the form of the cumulative number of craters, accumulated at the lunar surface per 1 Ga (Figure 12), shows that the ‘nonporous’ scaling law (eqn [8]) provides a good fit for large craters with diameters $D > 10 \text{ km}$. Production rate of small craters in regolith ($D < 300 \text{ m}$) is fitted well with the ‘porous rock’ (dry sand) scaling law (eqn [9]). In the intermediate size range nonporous scaling law results in the order of magnitude overestimated crater production rate near $D \sim 1 \text{ km}$. The data reconciliation is possible under assumption that upper lunar crust is porous enough in the upper $\sim 1 \text{ km}$ layer.

The assumption of validity of a ‘porous’ scaling law for lunar impacts, creating craters as large as $\sim 1 \text{ km}$ in diameter, looks unexpected at first glance. However, the inspection of published data on the geophysics from early publications (Toksoz *et al.*, 1972; Horvath *et al.*, 1980) to the recent data re-evaluation (Lognonné *et al.*, 2003) shows that all models assume the longitudinal seismic velocity as low as $\sim 1 \text{ km s}^{-1}$ above $\sim 1 \text{ km}$ depth below the lunar surface, and increase to $3\text{--}4 \text{ km s}^{-1}$ to a depth of $\sim 5 \text{ km}$. Comparing these estimates with experimental laboratory data for porous and shattered rocks one concludes that these low sound speeds cannot be attributed to fracturing only. The lithostatic pressure on the Moon at a depth of 1 km is about 4–5 MPa (40–50 bar), which is not enough to effectively close fracture porous space. Lunar basalt samples demonstrate void porosity of 3–30% and longitudinal sound speed of 3 km s^{-1} (Chung, 1972). Terrestrial rock samples of tuff and ignimbrites at 5 MPa load show longitudinal elastic wave velocities of 2.5 km s^{-1} having porosity of 40% (Vanorio *et al.*, 2002). The limited observations on clasts in lunar returned samples also demonstrate a range of porosity (at the hand sample level) from effective 0% to $\sim 40\%$ (Warren, 2001). Loose consolidation between successive basalt flows with regolith surfaces, and megaregolith gardening effects, may also affect the bulk mechanical property. Future investigations will answer the question whether real porosity in the upper 1 km of the lunar surface is large enough to prove the usage of the ‘porous rock’ scaling for small ($D < 10 \text{ km}$) lunar craters. Now we can state only that this assumption results in a good correspondence between the assumed current terrestrial bombardment rate and the lunar cratering rate, averaged on a wide historical time span (see Chapter 10.03).

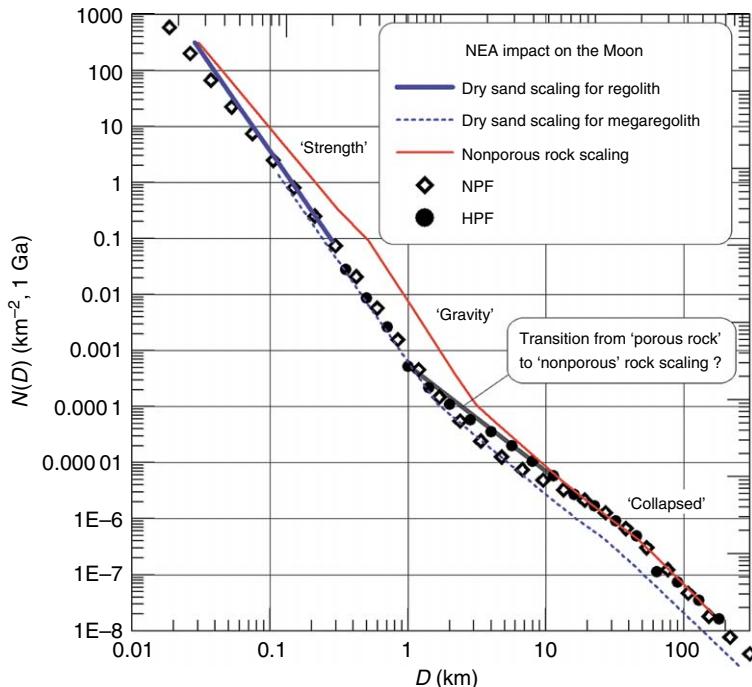


Figure 12 Comparison of the modern Earth/Moon projectile SFD with lunar production functions (NPF and HPF) in the form of the cumulative number of craters, accumulated at the lunar surface per 1 Ga, with different scaling laws. Usage of the ‘nonporous’ scaling law (eqn [8]) gives a good correspondence for crater diameters $D > 10$ km. Production of small crater in regolith ($D < 300$ m) is fitted well with the ‘porous rock’ scaling law (eqn [9]). In the intermediate size range nonporous scaling law results in a factor of 10 overestimate of the crater production rate near $D \sim 0.5$ km. The data reconciliation is possible under an assumption that upper lunar crust is porous enough in the upper ~ 1 km layer. See text for discussion. The factor of 10 discrepancy in the crater frequency for the same projectile SFD arises from the large assumed impact velocity on the Moon (~ 19 km s $^{-1}$), strongly decreasing the cratering efficiency for porous scaling law (see also [Figure 4](#)).

The outcome of this discussion for interplanetary surface dating is the possibility that on Mars and Mercury, where porosity of the upper subsurface rocks may be different due to different gravity and/or the presence of mechanisms of pore space infilling (e.g., water-caused weathering and massive amounts of ground ice at depths below few hundred meters), the scaling laws giving crater size D for a given meteoroid are different from that on the Moon, affecting the calculated crater production rate for a specified D .

10.06.6.3 Mercury

Using observed asteroid orbit distributions to estimate cratering rate relative to the Moon Neukum *et al.*, (2001b) find an average impact velocity of 32 km s $^{-1}$ and R_b of 0.7–0.85 relative to the Moon (for asteroids with $H < 17$ in [Table 4](#) $U \approx 35$ km s $^{-1}$ $R_b = 0.92$). Scaling laws (eqns [8] and [15]) result in the cratering rate ratio of 1.1 for $D < 1$ km and 0.9–1.3

for craters with $1 < D$ (km) < 100 . These results may be compared with an approach by Marchi *et al.* (2005) where the set of small body orbits is generated from the direct numeric modeling of orbital evolution of asteroids, initially placed near 3:1 and v_6 resonances in the MB. These authors estimate the impact probability on Mercury relative to Earth. Marchi *et al.* (2005) also take into account the fragmentation of small bodies due to collisions within the MB. The Mercury/Earth ‘bolide’ ratio is found to be $R_b \sim 1.08$ for large bodies ($D_p > 10$ m) and decreases for smaller projectiles due to progressive fragmentation (the ratio of total number of impacts over all of the planetary surface is 0.16), the mean impact velocity is about 30 km s $^{-1}$ (the impact velocity covers the range from ~ 15 to ~ 80 km s $^{-1}$). Assuming the Earth/Moon bolide ratio of the order of 1.68 ([Table 4](#)), we estimate $R_b \sim 1.8$ for Mercury relative to the Moon. This value is twice larger than that derived by Neukum *et al.* (2001b), illustrating the current accuracy of interplanetary impact rate comparisons. One should

note that Marchi *et al.* (2005) model projectiles originated near most powerful resonances in the MB, while a considerable NEA subpopulation ($\sim 1/3$ of all asteroid-like objects) is resupplying via a ‘slow track’ from the Mars-crossers reservoir. The inclusion of the Mars-crosser source of NEA into modeling may change the modeling result to some extent.

10.06.6.4 Venus

According to simple estimates (Table 4), the R_b Venus/Moon is close to 1.1 from observed NEA with $H < 17$, giving the impact rate ~ 1.5 times smaller than on Earth but larger average impact velocity. The approximation of nearly the same cratering rate on Earth and Venus has resulted in the first average age estimates, based on radar images from the Russian Venera 15/16 mission (Ivanov *et al.*, 1986; Ivanov and Bazilevskii, 1987). From early data processing Hartmann (1987) have published lunar-based crater SFD isochrons for Venus. McKinnon *et al.* (1997) have used small body orbital set derived from earlier observations (Shoemaker *et al.*, 1990, 1994) to derive the average crater retention age of Venusian surface as ~ 700 Ma.

10.06.6.5 Mars

The crater count chronology technique appears to have had a good predictive success as applied to Mars. Starting with Mariner 9 data, Hartmann used the larger Martian craters ($D > \sim 500$ m) and crude early scaling techniques in a series of papers (Hartmann, 1973; Hartman, 1978; Hartmann *et al.*, 1981); consistently deriving ages for the surface flows in broad Amazonian lava plains and major volcanic constructs such as Olympus Mons of the order 100–2000 Ma. These ages were questioned at the time by some workers who thought they might be too low. In the 1980s, it was recognized that Martian meteorites are mostly lavas and other igneous rocks in the age range 170–1300 Ma. This consistency gives some confidence as to the general reliability of the methods.

In recent years, with Mars orbiter camera (MOC, NASA Global Surveyor mission), thermal emission imaging system (THEMIS, NASA Mars Odyssey mission) and high-resolution stereo camera (HRSC) cameras returning images showing craters as small as $D = 10$ m, on various geologic formations of various scales, there has been increasing emphasis on the question of whether the small craters can be used to date small-scale formations of great interest, such as

glacier-like features, debris aprons, floors of fluvial channels, landslide deposits, mantled deposits, and exhumed surfaces. Ages as low as a few million years ago, have been proposed from crater counts on many of these features.

Figures 13(a) and 13(b) give few examples of the real crater counts and their correspondence to model isochrones (Hartmann, 2005). Crater counts averaging, using Mariner and Viking data and an average of several Mars Global Surveyor (MGS) and THEMIS frames, suggesting that the surface of lavas average no more than a few hundred million years ago in age. Crater counts in and around various parts of Marte Vallis indicate that this system is much younger than most Martian outflow channel systems.

The decades of Mars imaging with orbital space-crafts accumulate very interesting data about impact cratering rate in the last ~ 10 years. The repetition of imaging of the same areas year by year now cover enough large surface to find impact craters formed in between two consequent imaging. Malin *et al.* (2006) have collected information about 20 new craters, appeared in $21 \times 10^6 \text{ km}^2$ of Martian surface between May 1999 and March 2006. Some of these 20 craters with diameters from 10 to ~ 150 m are clustered impacts, most probably due to splitting of a projectile in the Marian atmosphere. Reduced to the equivalent single crater, these impacts may be binned and compared with Hartmann’s isochrone (Figure 13(c)). Except the largest crater, 19 other craters well fit the strip between 1 and 10 years isochrones. It gives some additional confidence to the whole system of terrestrial planet’s surface age comparison with the cratering statistics. However, for surfaces older than ~ 1 Ma small craters may be produced not only by meteoroids from heliocentric orbits, but also with fragments, ejected from large ($> 3\text{--}10$ km) impact craters. The 1 Ma age is mentioned as currently satellite imaging has not yet revealed a young crater of this size. However, for older areas the crater chronology is now under discussion in respect of usage of crater counts below crater diameters ~ 100 m. In a section below, we take up some of the problems of applying small craters.

10.06.6.6 Problem of Interpretation of Small Craters

The basic principle of crater dating of planetary surfaces is the assumption that older surfaces accumulate more craters per unit area than young surfaces. This principle may be confused by many processes such as erosion and deposition, making the

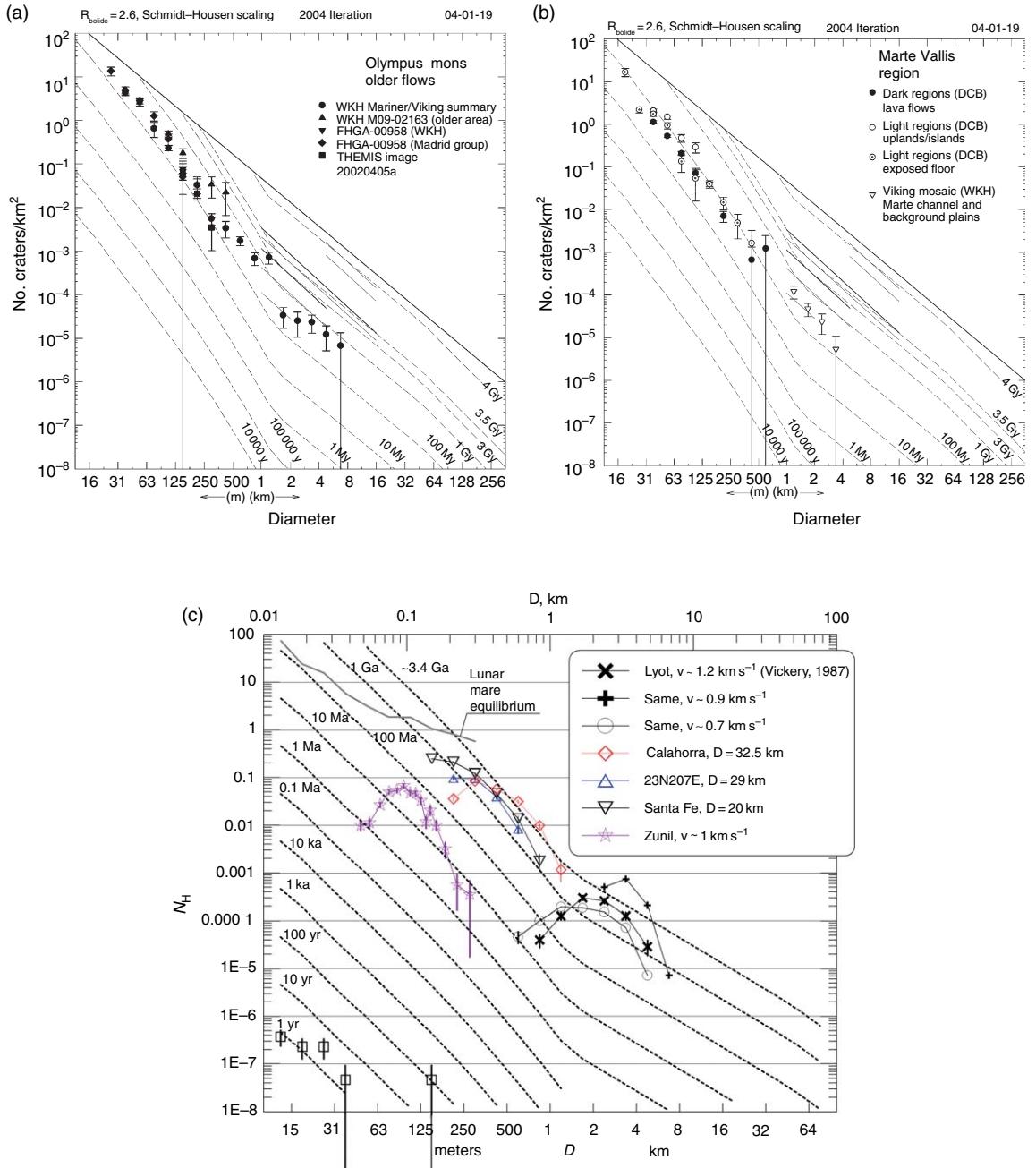


Figure 13 (a) Crater counts averaging over older background slope surfaces of Olympus Mons, using Mariner and Viking data and an average of several MGS and THEMIS frames, suggesting that the upper few hundred meters of near-surface lavas average no more than a few hundred million years ago in age (Hartmann, 2005). (b) Marte Vallis, a fluvial system cut into recent lava plains, is distinguished in this region by a dark-toned lava flow that entered the channel floor. Crater counts in and around various parts of Marte Vallis fit an isochron with model age around 30–300 Ma, indicating that this system is much younger than most Martian outflow channel systems. (c) Secondary crater counts (Ivanov, 2006; Werner et al., 2006) in Hartmann's coordinates show an appreciable difference due to a bell-shaped behavior. The same plot is used to show the first direct estimates of small impact crater production rate on Mars (open squares with errorbars), revealed from images, obtained between May 1999 and March 2006 (Malin et al., 2006), in comparison with Hartmann's isochrones for 1 and 10 years of small crater accumulation. Parts (a) and (b) are reprinted from Hartmann WK (2005) Martian cratering 8: Isochron refinement and the chronology of Mars. *Icarus* 174: 294–320, with permission from Elsevier.

dating (both absolute and relative) ambiguous in a specific situation. The ratio of crater formation rate from celestial small bodies (primary impacts) to cratering from ‘secondary’ fragments ejected from primary craters is uncertain. The gradual accumulation of small primary craters may be overwhelmed by sporadic showers of secondaries from remote large primaries, adding another complication to estimating the crater retention age. In the system of Hartmann, the crater counts avoid obvious clusters and rays, in hopes of measuring the global accumulation of primaries (at small sizes) and dearly random ‘distant secondaries’. Distant secondaries are those accumulating around the planet, far from their parent primary. Discussing the effect of distant secondary craters, one should accept a few main points: (1) The amount of material and, consequently number of fragments, creating secondary craters, is limited by the total amount of primary crater ejecta. (2) The appearance of secondary craters on different planets may be quite different depending on the presence of atmosphere.

Secondary craters on Earth are known mainly for two kinds of processes: volcanic eruptions and explosion cratering. Volcanic secondary craters are observed typically for terminal stages of volcanic eruptions, when volcanic bombs are ejected and land on a nearby smooth, for example, a surface covered with volcanic ash (Hartmann, 1967). Secondary craters around large-scale chemical and nuclear detonation craters have been discussed by Roberts (1964). For largest excavation events those secondary craters reach diameters of ~ 40 m. A few reports about secondary craters around natural terrestrial impact craters are not convincing enough leaving room for further studies (Poag, 1995, 1999). In all cases secondary craters have been found not far from the source and assumed velocity of ejection are in the range of $n \times 100$ m s $^{-1}$ and below.

The recognition of ray craters on the lunar surface inspired investigations both of impact nature of lunar craters and recognition of secondary craters, especially along visible crater rays (Fielder, 1962). The first close look (from Ranger spacecraft) revealed small craters, invisible from Earth, and dramatically increased the understanding of impact cratering as the main landscape-formation process on the Moon (Kuiper and Hartmann, 1965; Shoemaker, 1965). The idea that majority of small craters are secondary craters was proposed immediately (Shoemaker, 1965). At the same time the idea was criticized from the point of view that material accelerated to high

velocity is not presented as single blocks of sufficient size (Kopal, 1966).

To accelerate such a rock (~ 270 m in diameter – BAI) to a velocity of 1.4 km s $^{-1}$ and transport in free flight to a place 1000 km distant from its origin in one piece would constitute a rather formidable requirement. Fortunately, it appears to us now that nature may no longer have to be held responsible for this task.

Issues were raised about the ratio of secondaries to primaries among different sized craters, but it was usually believed that the ratio is low in the ‘shallow branch’ of the lunar SFD, at $D > 2$ km.

10.06.6.6.1 Secondary craters on Mars

Finding of remote secondary crater fields on Mars and Europa (Bierhaus *et al.*, 2001, 2005; McEwen *et al.*, 2005) revitalized the old discussion about the proportion of secondary and primary impact craters in crater counts used to date planetary surfaces. The existence of small ‘background’ secondary craters is not in doubt as we have lunar and Martian meteorites: if an impact produces sizable escaping ejecta fragments, the appreciable amount of fragments with sub-escape velocity should create globally scattered secondary craters. A general review of ideas and findings is compiled by McEwen and Bierhaus (2006). The main point of McEwen and Bierhaus is that after one large primary crater creates a lot of remote secondary craters, secondary craters dominate in crater records at small crater diameters on surfaces older than the age of that primary. They also propose that the nonuniform spatial distribution of distant secondaries around a planet makes questionable crater retention age determination based on small crater counts.

McEwen *et al.* (2005), Bierhaus, and McEwen and Bierhaus (2006) have all stated as a premise that crater counting techniques fundamentally depend on an assumption that all craters are primaries. For example, McEwen *et al.* (2005) cite numbers from the Hartmann–Neukum production rates and the isochrons and then refer to them as overestimating the number of primaries by factors as much as 700–2000. This is seriously misleading, because the Hartmann isochrons include not just primaries but the total mix of primaries plus distant secondaries. The counts include all the craters, excluding obvious clusters and rays, in hopes of gaining chronologic information from the global buildup of the full SFD. In proof of

this, McEwen *et al.* (2005) give a new discussion of the age of Athabasca Vallis; instead of revising earlier estimates by a factor 2000, they give exactly the same age range (a few to ~ 200 Ma) as had been published few years before by Berman and Hartmann. Werner *et al.* (2003) came to similar results independently. For these reasons we believed much of the early discussion of ‘the problem of secondaries’ was overheated (Werner *et al.*, 2003).

McEwen *et al.* (2005; table 3) give predicted SFDs of secondaries at different distances from Zunil-sized craters. Based on these numbers, one can show that a single Zunil (forming about 1/Ma on Mars) produces 20 m scale craters about equivalent of Hartmann’s 1 Ma isochron over 1/6 of Mars. Thus after something like 6–20 Ma (taking into account that Mars uplands produce fewer secondaries than Amazonian lavas), one can show that most of Mars would have 20 m crater densities comparable to Hartmann’s isochrons for a few My. In about 100 Ma the secondaries are asymptotically approaching the 100 My isochron. This suggests that even the smallest secondaries do carry chronological information. Models with 20 m scale craters being mostly primaries, or the McEwen *et al.* model where they are mostly secondaries, either way, predict that a surface larger than a few kilometers across with no visible 20 m craters is probably less than a few million years or few tens of million years old. Interestingly, such surfaces would be formed with the last one or few high-obliquity episodes of climate change and dust/ice mantling. Thus, they may provide information about climatic and mantling cycles.

We have not yet dated returned samples from Mars, so arguments pro and contra in the discussion about error bars in crater retention age are mostly based on models. The set of observation data should be added to the discussion. Implicitly, McEwen and Bierhaus assume that remote secondaries from each primary crater cannot be recognized without additional high-resolution images and analysis of individual shapes of small craters.

10.06.6.6.2 SFD for secondary craters

Shoemaker (1965) proposed that secondary craters have a very steep slope SFD: in cumulative form $N \sim D^{-m}$, and $m \approx 4$ (in contrast to $m \leq 3$ for small primary craters). The inspection of two examples for secondary craters, used by Shoemaker, shows that $N(D)$ curves are concaved and the local exponent m is slowly decreased when crater diameter decreases. This is a typical behavior for SFD of rock fragments,

produced by an impulsive rock demolition. Decades of engineering experience resulted in many forms of fragment SFD description, and one of the most usable SFD forms is the so-called Weibull distribution (Weibull, 1951), reviewed in detail by Grady and Kipp (1987) in application for explosive rock fragmentation. The Weibull SFD in its cumulative form expresses the volume of fragments larger than a given size in a fragment population analyzed:

$$V(>x) = V_0 \exp [-(x/x^*)^n] \quad [27]$$

where V_0 is the total volume of fragments, x^* expresses the ‘characteristic size’ in a population of fragments, and n is the so-called Weibull exponent. According to Grady and Kipp (1987) the Weibull exponent varies from ~ 6 for a few fragmentation acts to ~ 1 for heavily comminuting processes (like a long-term fragmentation in rock mills). For explosion fragments n are in the range from 1 to 3 typically. The characteristic feature of the Weibull SFD is the ‘rollover’ at smaller sizes – the distribution for fine fragments is much more shallow than for largest fragments. This may be valid for fragments ejected from a crater and created secondary craters. Hence, smaller members in a cluster of secondary craters may not be so numerous as is widely supposed from the steep slope for larger members. The rollover character of SFD in clusters of recognizable shape and ejecta pattern lunar and Martian secondary craters is demonstrated by Ivanov (2006).

The available SFDs for close and remote secondary craters on Mars are shown in Figure 13(c) including the key case of Zunil. As McEwen’s numerical data for SFD of secondaries are unavailable to date, we use HRSC crater count (Werner *et al.*, 2006) in the area ~ 300 km west of Zunil (with the corresponding ejection/landing fragment velocity of about 1 km s^{-1}). It is interesting that fragments that created secondaries around much larger Lyot crater in the same velocity range also have a ‘bell-shaped’ SFD (Ivanov, 2006). According to both McEwen *et al.* (2005) and the HPF isochrones (Hartmann, 2005), one Zunil-size primary crater should form on Mars every 1 Ma, on average. Consistent with this, the secondary SFD measured in the HRSC image is well bracketed with 1 and 10 Ma model isochrons. The secondary crater densities rise above the 1 Ma isochron at this distance from Zunil, but at larger distances they drop below (according to McEwen’s model). These facts suggest that the HPF isochrones lie closer to Martian reality on surfaces as young as a few Ma, than suggest by McEwen *et al.*

(2005). A better test, of course, is to try to take into account the total accumulation rate of secondaries from all sizes of primaries (not just Zunil-size craters). We note, however, that primary craters of $D < 3$ km are too small to launch large distance secondaries (Head *et al.*, 2002).

Available published data on SFD of secondary craters and ejected blocks demonstrate that individual close and remote secondary crater fields may be fitted with the bell-shaped SFD (like the Weibull statistics) standard for fragmentation problems. For an individual primary crater secondary craters at a certain distance from a parent crater are created with fragments, ejected with the same velocity and ejection angle. Hence, if secondary craters from one primary crater are much more numerous than primary craters at the given location, they are easily recognizable by the bell-shaped (in R -plot) Weibull-like SFD in addition to clustering of secondaries. Thus we infer that many primary craters should be accumulated at a planet globally before the summation of secondary craters from multiple primaries would reproduce the primary SFD shape. The question of how many unrecognized secondaries are incorporated into crater counts used for surface dating definitely should be investigated quantitatively in future.

These speculations, together with the steep SFD slope of impact craters on Gaspra, suggest that the interplanetary flux of primaries may be the significant contribution to the ‘steep branch’ of SFD for small craters observed on the Moon, Mars, and Gaspra. More data about small craters on asteroids show a complicated game of production/erosion of small craters on asteroids (Chapman, 2002; Chapman *et al.*, 2002; Richardson *et al.*, 2005).

More work should be done to clarify the real level of inaccuracy of surface dating due to existence of ‘background’ secondary craters.

10.06.7 Conclusions

The exogenic dynamics of planets is a fast developing system of knowledge, integrating facts and models from many fields of science from celestial mechanics to physics of shock waves. Visible on all planetary bodies the scars of impact – impact craters – vividly illustrate that exchange in solid bodies is an active agent of the solar system and planetary surface evolution. Individual giant impacts appreciably change the heat balance of planets, producing ‘impact hot spots’ for post-impact magmatic activity. SFDs of numerous

small impacts, recorded in crater populations, create an opportunity not only to estimate the age of planetary surfaces, but also the rates and efficiency of erosion and deposition processes (through studies of degradation and losses of small craters). SDF of craters can be compared with SDF of asteroids and comets, resulting in observational constraints for the general solar system evolution picture. Many topics of exogenic dynamics are still under construction, and many problems so far have only qualitative solutions. Progress in the exogenic dynamics offer better understanding of planetary formation and evolution.

References

- Andre SL and Watters TR (2006) Depth to diameter measurements of Mercurian mature complex craters. *LPSC 37th*: #2054.
- Andre SL, Watters TR, and Robinson MS (2005) The long wavelength topography of Beethoven and Tolstoj basins, Mercury. *Geophysical Research Letters* 32: 21202.
- Arnold JR (1965) The origin of meteorites as small bodies. Part II: The model. *Astrophysical Journal* 141: 1536–1547.
- Artemieva NA and Ivanov BA (2001) Numerical simulation of oblique impacts: Impact melt and transient cavity size. *Lunar and Planetary Science Conference* 32: #1321.
- Arthur DWG, Agnieray AP, Horvath RA, Wood CA, and Chapman CR (1963) The system of lunar craters, quadrant I. *Communications of the Lunar and Planetary Laboratory* 3: 30–60.
- Arthur DWG, Agnieray AP, Horvath RA, Wood CA, and Chapman CR (1965a) The system of lunar craters, quadrant II. *Communications of the Lunar and Planetary Laboratory* 3: 1–60.
- Arthur DWG, Agnieray AP, Pellicori RA, Wood CA, and Weller T (1965b) The system of lunar craters, quadrant III. *Communications of the Lunar and Planetary Laboratory* 3: 61–146.
- Arthur DWG, Pellicori RA, and Wood CA (1966) The system of lunar craters, quadrant IV. *Communications of the Lunar and Planetary Laboratory* 5: 1–208.
- Bandermann LW and Singer SF (1973) Calculation of meteoroid impacts on Moon and Earth. *Icarus* 19: 108–113.
- Bertoldi F, Altenhoff W, Weiss A, Menten KM, and Thum C (2006) The trans-neptunian object UB313 is larger than Pluto. *Nature* 439: 563–564.
- Bierhaus EB, Chapman CR, and Merline WJ (2005) Secondary craters on Europa and implications for cratered surfaces. *Nature* 437: 1125–1127.
- Bierhaus EB, Chapman CR, Merline WJ, Brooks SM, and Asphaug E (2001) Pwyll secondaries and other small craters on Europa. *Icarus* 153: 264–276.
- Binzel RP, Rivkin AS, Stuart JS, Harris AW, Bus SJ, and Burbine TH (2004) Observed spectral properties of near-Earth objects: Results for population distribution, source regions, and space weathering processes. *Icarus* 170: 259–294.
- Borovička J, Koten P, Spurný P, Boček J, and Stork R (2005) A survey of meteor spectra and orbits: evidence for three populations of Na-free meteoroids. *Icarus* 174: 15–30.
- Bottke J, William F, Durda DD, *et al.* (2005) Linking the collisional history of the main asteroid belt to its dynamical excitation and depletion. *Icarus* 179: 63–94.

- Bottke WF and Greenberg R (1993) Asteroidal collision probabilities. *Geophysical Research Letters* 20: 879–881.
- Bottke J, William F, Morbidelli A, et al. (2002) Debiasing orbital and absolute magnitude distribution of the near-Earth objects. *Icarus* 156: 399–433.
- Bottke J, William F, Vokrouhlický D, Rubincam DP, and Nesvorný D (2006a) The Yarkovsky and YORP effects: Implications for asteroiddynamics. *Annual Review of Earth and Planetary Sciences* 32: 157–191.
- Bottke J, William F, Vokrouhlický D, Rubincam DP, and Nesvorný D (2006b) The Yarkovsky and YORP effects: Implications for asteroid dynamics. *Annual Review of Earth and Planetary Sciences* 32: 157–191.
- Brett D, Yeomans TD, Housen K, and Consolmagno G (2002) Asteroid density, porosity, and structure. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 485–500. Tucson, AZ: University of Arizona Press.
- Brown P, Spalding RE, ReVelle DO, Tagliaferri E, and Worden SP (2002) The flux of small near-Earth objects colliding with the Earth. *Nature* 420: 294–296.
- Burchell MJ and Whitehorn L (2003) Oblique incidence hypervelocity impacts on rock. *Monthly Notices of the Royal Astronomical Society* 341: 192–198.
- Capek D and Vokrouhlický D (2004) The YORP effect with finite thermal conductivity. *Icarus* 172: 526–536.
- Cepeláček Z (1987) Geometric, dynamic, orbital and photometric data on meteoroids from photographic fireball networks. *Bulletin of the Astronomical Institutes of Czechoslovakia* 38: 222–234.
- Chambers JE and Wetherill GW (2001) Planets in the asteroid belt. *Meteoritics and Planetary Science* 36: 381–399.
- Chapman CR (2002) Cratering on asteroids from Galileo and NEAR Shoemaker. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 315–329. Tucson, AZ: University of Arizona Press.
- Chapman CR and Haefner RR (1967) A critique of methods for analysis of the diameter-frequency relation for craters with special application to the Moon. *Journal of Geophysical Research* 72: 549.
- Chapman CR, Merline WJ, Thomas PC, Joseph J, Cheng AF, and Izenberg N (2002) Impact history of Eros: Craters and boulders. *Icarus* 155: 104–118.
- Chesley SR, Ostro SJ, Vokrouhlický D, et al. (2003) Direct detection of the Yarkovsky effect by radar ranging to asteroid 6489 Golevka. *Science* 302: 1739–1742.
- Chung DH (1972) Laboratory studies on seismic and electrical properties of the Moon. *Moon* 4: 356.
- Cintala MJ and Grieve RAF (1998a) Erratum: Scaling impact melting and crater dimensions: Implications for the lunar cratering record. *Meteoritics and Planetary Science* 33: 1343.
- Cintala MJ and Grieve RAF (1998b) Scaling impact-melt and crater dimensions: Implications for the lunar cratering record. *Meteoritics and Planetary Science* 33: 889–912.
- Collins GS, Melosh HJ, Morgan JV, and Warner MR (2002) Hydrocode simulations of Chicxulub Crater collapse and peak-ring formation. *Icarus* 157: 24–33.
- Croft SK (1981) The excavation stage of basin formation: A qualitative model. In: Schultz PH and Merrill RB (eds.) *Multiring Basins*, pp. 207–225. New York: Pergamon Press.
- Croft SK (1985) The scaling of complex craters. Proceedings of the 15th Lunar and Planetary Science Conference. *Journal of Geophysical Research* 90: C828–C842.
- Crosby A and McKenzie D (2005) Measurements of the elastic thickness under ancient lunar terrain. *Icarus* 173: 100–107.
- D'Abramo G, Harris AW, Boattini A, Werner SC, Harris AW, and Valsecchi GB (2001) A simple probabilistic model to estimate the population of near-Earth asteroids. *Icarus* 153: 214–217.
- Dence MR (1965) The extraterrestrial origin of Canadian craters. *Annals of the New York Academy of Sciences* 123: 941–969.
- Dienes JK and Walsh JM (1970) Theory of impact: Some general principles and the method of Eulerian codes. In: Kinslow R (ed.) *High Velocity Impact Phenomena*, pp. 45–104. New York: Academic Press.
- Dones L, Gladman B, Melosh HJ, Tonks WB, Levison HF, and Duncan M (1999) Dynamical lifetimes and final fates of small bodies: Orbit integrations vs Öpik calculations. *Icarus* 142: 509–524.
- Farinella P, Froeschlé C, Froeschlé C, et al. (1994) Asteroids falling onto the Sun. *Nature* 371: 315–317.
- Fernandez JA, Gallardo T, and Brunini A (2002) Are there many inactive Jupiter-family comets among the near-Earth asteroid population? *Icarus* 159: 358–368.
- Fielder G (1962) Ray elements and secondary-impact craters on the Moon. *Astrophysical Journal* 135: 632–637.
- French BM and Short NM (eds.) (1968) *Shock Metamorphism of Natural Materials*, 644 pp. Baltimore: Mono Book Corporation.
- Fujiiwara A, Kawaguchi J, Uesugi K, et al. (2006) Global properties of 25143 Itokawa observed by Hayabusa. *Lunar and Planetary Science Conference 37th* 37: #1575.
- Gallant J and Gladman B (2006) Lunar cratering asymmetries. *Lunar and Planetary Science Conference 37th* abstract no 2336.
- Gault DE and Wedekind JA (1978) Experimental studies of oblique impact. In: Lunar Science Institute (ed.) *Proceedings of the Lunar and Planetary Science Conference 9th*, pp. 3843–3875 pp. New York: Pergamon Press.
- Gerasimov MV, Ivanov BA, Yakovlev Ol, and Dikov YP (1999) Physics and chemistry of impacts. In: Ehrenfreund P, Kraft C, Kochan H, and Pirronello V (eds.) *Laboratory Astrophysics and Space Research*, pp. 209–259. Boston, MA: Kluwer Academic Publishers.
- Gilbert GK (1893) The Moon's face. A study of the origin of its features. *Philosophical Society of Washington, Bulletin VII*: 241–292.
- Gladman BJ, Burns JA, Duncan MJ, and Levison HF (1995) The dynamical evolution of lunar impact ejecta. *Icarus* 118: 302–321.
- Gladman B, Michel P, and Froeschlé C (2000) The near-Earth object population. *Icarus* 146: 176–189.
- Gomes R, Levison HF, Tsiganis K, and Morbidelli A (2005) Origin of the cataclysmic late heavy bombardment period of the terrestrial planets. *Science* 335: 466–469.
- Grady DE and Kipp ME (1987) Dynamic rock fragmentation. In: Atkinson BK (ed.) *Fracture Mechanics of Rock*, pp. 429–475. San Diego, CA: Academic.
- Green DWE (2006) (134340) Pluto, (136199) Eris, and (136199) Eris I (Dysnomia). *International Astronomical Union Circular* 8747: 1.
- Greenberg R and Hartmann WK (1977) Impact strength: A fundamental parameter of collisional evolution (abstract). *Bulletin of the American Astronomical Society* 9: 455.
- Greenberg R, Hartmann WK, Chapman CR, and Wacker JF (1978) The accretion of planets from planetesimals. In: Gehrels T and Matthews MS (eds.) *Protostars and Planets: Studies of Star Formation and of the Origin of the Solar System*, pp. 599–620. Tucson, AZ: University of Arizona Press.
- Grieve RA (1984) The impact cratering rate in recent time. *Journal of Geophysical Research Supplement* 89: 403.
- Grieve RAF and Cintala MJ (1992) An analysis of differential impact melt-crater scaling and implications for the terrestrial impact record. *Meteoritics* 27: 526–538.
- Grieve RAF, Dence MR, and Robertson PB (1977) Cratering processes: As interpreted from the occurrence of impact melts. In: Roddy DJ, Pepin RO, and Merrill RB (eds.) *Impact and Explosion Cratering*, pp. 791–814. New York: Pergamon Press.
- Grieve RAF and Garvin JB (1984) A geometric model for excavation and modification at terrestrial simple craters. *Journal of Geophysical Research* 89: 11561–11572.

- Grieve RAF and Pesonen LJ (1995) Terrestrial impact craters: Their spatial and temporal distribution and impacting bodies. *Earth, Moon, and Planets* 71: 357–376.
- Grieve RAF and Shoemaker EM (1994) The record of the past impacts on Earth. In: Gehrels T (ed.) *Hazards Due to Comets and Asteroids*, pp. 417–462. Tucson, AZ: University of Arizona Press.
- Halliday I, Griffin AA, and Blackwell AT (1996) Detailed data for 259 fireballs from the Canadian camera network and inferences concerning the influx of large meteoroids. *Meteoritics and Planetary Science* 31: 185–217.
- Harris AW (2003) The impact frequency of near-Earth asteroids. *AAS/Division of Dynamical Astronomy Meeting* 34: abstract #02.04.
- Hartmann WK (1965) *History of the Lunar Surface*, 179 pp. Tucson, AZ: University of Arizona.
- Hartmann WK (1966a) Martian cratering. *Icarus* 5: 565–576.
- Hartmann WK (1966b) Early lunar cratering. *Icarus* 5: 565–576.
- Hartmann WK (1967) Secondary volcanic impact craters at Kapoho, Hawaii and comparisons with the lunar surface. *Icarus* 7: 66–75.
- Hartmann WK (1969) Terrestrial, lunar, and interplanetary rock fragmentation. *Icarus* 10: 201–213.
- Hartmann WK (1970a) Note: Lunar cratering chronology. *Icarus* 13: 299.
- Hartmann WK (1970b) Note: Preliminary note on lunar cratering rates and absolute time-scales. *Icarus* 12: 131.
- Hartmann WK (1972) Interplanet variations in scale of crater morphology Earth, Mars, Moon. *Icarus* 17: 707.
- Hartmann WK (1973) Martian cratering 4: Mariner 9 initial analysis of cratering chronology. *Journal of Geophysical Research* 78: 4096–4116.
- Hartmann WK (1977) Relative crater production rates on planets. *Icarus* 31: 260–276.
- Hartman WK (1978) Martian cratering V – Toward an empirical Martian chronology, and its implications. *Geophysical Research Letters* 5: 450–452.
- Hartmann WK (1987) Determination of crater retention ages on Venus: A methodology. *Izvestia Akademii Nauk SSSR. Seria Geologicheskaya* 1987: 67–74 (Article in Russian).
- Hartmann WK (2005) Martian cratering 8: Isochron refinement and the chronology of Mars. *Icarus* 174: 294–320.
- Hartmann WK, Anguita J, de la Casa MA, Berman DC, and Ryan EV (2001) Martian cratering. Part 7: The role of impact gardening. *Icarus* 149: 37–53.
- Hartmann WK and Davis DR (1975) Satellite-sized planetesimals and lunar origin. *Icarus* 24: 504–514.
- Hartmann WK, Farinella P, Vokrouhlický D, et al. (1999) Reviewing the Yarkovsky effect: New light on the delivery of stone and iron meteorites from the asteroid belt. *Meteoritics & Planetary Science* 34: 161–167.
- Hartmann WK and Kuiper GP (1962) Concentric structures surrounding lunar basins. *Communications of the Lunar and Planetary Laboratory, University of Arizona* 1: 51–66.
- Hartmann WK and Neukam G (2001) Cratering chronology and evolution of Mars. In: Kallenbach R, Griss J, and Hartmann WK (eds.) *Chronology and Evolution of Mars*, pp. 165–194. Dordrecht: Kluwer Academic Press.
- Hartmann WK, Quantin C, and Mangold N (2007) Possible long-term decline in impact rates. Part 2: Lunar impact-melt data regarding impact history. *Icarus* 186: 11–23.
- Hartmann WK, Ryder G, Dones L, and Grinspoon D (2002) The time-dependent intense bombardment of the primordial Earth/Moon system. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 493–512. Tucson, AZ: University of Arizona Press.
- Hartmann WK, Strom RG, Weidenschilling SJ, et al. (1981) Chronology of planetary volcanism by comparative studies of planetary craters. In: Houston Lunar and Planetary Institute (eds.) *Basaltic Volcanism on the Terrestrial Planets*, pp. 1050–1127. Elmsford., NY: Pergamon Press.
- Hartmann WK and Wood CA (1971) Moon: Origin and evolution of multiring basins. *Moon* 3: 3–78.
- Head JN, Melosh HJ, and Ivanov BA (2002) Martian meteorite launch: High-speed ejecta from small craters. *Science* 298: 1752–1756.
- Hiesinger H and Head JW, III (2002) Topography and morphology of the Argyre Basin, Mars: Implications for its geologic and hydrologic history. *Planetary and Space Science* 50: 939–981.
- Hiesinger H, Head JW, Wolf U, Jaumann R, and Neukam G (2003) Ages and stratigraphy of mare basalts in Oceanus Procellarum, Mare Nubium, Mare Cognitum, and Mare Insularum. *Journal of Geophysical Research* 108: 5065 (doi:10.1029/2002JE001985).
- Hiesinger H, Jaumann R, Neukam G, and Head JW (2000) Ages of mare basalts on the lunar nearside. *Journal of Geophysical Research* 105: 29239–29276.
- Holsapple KA (1993) The scaling of impact processes in planetary sciences. *Annual Review of Earth and Planetary Sciences* 21: 333–373.
- Holsapple KA and Schmidt RM (1979) A material-strength model for apparent crater volume. In: Lunar Science Institute (ed.) *Proceedings of the 10th Lunar and Planetary Science Conference*, pp. 2757–2777. New York: Pergamon.
- Holsapple KA and Schmidt RM (1980) On the scaling of crater dimensions. Part 1: Explosive processes. *Journal of Geophysical Research* 85: 7247–7256.
- Holsapple KA and Schmidt RM (1982) On the scaling of crater dimensions. Part II: Impact processes. *Journal of Geophysical Research* 87: 1849–1870.
- Holsapple KA and Schmidt RM (1987) Point source solutions and coupling parameters in cratering mechanics. *Journal of Geophysical Research* 92: 6350–6376.
- Horvath P, Latham GV, Nakamura Y, and Dorman HJ (1980) Lunar near-surface shear wave velocities at the Apollo landing sites as inferred from spectral amplitude ratios. *Journal of Geophysical Research* 85: 6572–6578.
- Housen KR and Holsapple KA (2003) Impact cratering on porous asteroids. *Icarus* 163: 102–119.
- Hughes DW (2000) A new approach to the calculation of the cratering rate of the Earth over the Last 125 ± 20 Myr. *Monthly Notices of the Royal Astronomical Society* 317: 429–437.
- Ivanov BA (1981) Cratering mechanics. In: *Advances in Science and Technology of VINITI, Ser. Mechanics of Deformable Solids*, pp. 60–128. Moscow: VINITI Press, (English translation: NASA Tech. Memorandum 88477/N87-15662, 1986).
- Ivanov BA (2001) Mars/Moon cratering rate ratio estimates. In: Kallenbach R, Griss J, and Hartmann WK (eds.) *Chronology and Evolution of Mars*, pp. 87–104. Dordrecht: Kluwer Academic Press.
- Ivanov BA (2004) Heating of the Lithosphere during Meteorite Cratering. *Solar System Research* 38: 266–279.
- Ivanov BA (2005) Numerical modeling of the largest terrestrial meteorite craters. *Solar System Research* 39: 381–409.
- Ivanov BA (2006) *Notes About Secondary Crater SFD*. Workshop on Surface Ages and Histories: Issues in Planetary Chronology, 23–23 May, 2006, LPI Contribution No. 1320, pp. 28–29. Huston, TX: Lunar and Planetary Institute Houston.
- Ivanov BA and Bazilevskii AT (1987) Comparison of crater retention ages on Earth and Venus. *Astronomicheskii Vestnik* 21: 136–143.
- Ivanov BA and Artemieva NA (2001) Transient cavity scaling for oblique impacts. *Lunar and Planetary Science Conference* 32: #1327.

- Ivanov BA and Artemieva NA (2002) Numerical modeling of the formation of large impact craters. In: Koeberl C and MacLeod KG (eds.) *Geological Society of America Special Papers 356: Catastrophic Events and Mass Extinctions: Impact and Beyond*, pp. 619–630. Boulder, CO: GSA.
- Ivanov BA, Bazilevskii AT, Krivchikov VP, and Cherniai IM (1986) Impact craters of Venus – Analysis of Venera 15 and 16 data. *Journal of Geophysical Research* 91: 413.
- Ivanov BA, Nemchinov IV, Svetsov VA, Provalov AA, Khazins VM, and Phillips RJ (1992) Impact cratering on Venus: Physical and mechanical models. *Journal of Geophysical Research* 97: 16167–16181.
- Ivanov BA, Neukum G, Bottke WF, Jr., and Hartmann WK (2002) The comparison of size-frequency distributions of impact craters and asteroids and the planetary cratering rate. *Asteroids III*: 89–101.
- Ivanov BA, Neukum G, and Wagner R (2001) Size-frequency distributions of planetary impact craters and asteroids. In: Marov MY and Rickman H (eds.) *Collisional Processes in the Solar System*, pp. 1–34. Dordrecht: Kluwer Academic Publishers.
- Ivanov BA and Stöffler D (2005) The Steinheim impact crater, Germany: Modeling of a complex crater with central uplift. *LPSC 36th*: abstract #1443.
- Ivezic Z, Tabachnik S, Rafikov R, et al. (2001) Solar system objects observed in the Sloan Digital Sky Survey Commissioning Data. *Astronomical Journal* 122: 2749–2784.
- Jedicke R, Larsen J, and Spahr T (2002) Observational selection effects in asteroid surveys. In: Bottke WF, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 71–87. Tucson, AZ: University of Arizona Press.
- Joeleht A, Kirsmae K, Plado J, Versh E, and Ivanov B (2005) Cooling of the Kardla impact crater. Part II: Impact and geothermal modeling. *Meteoritics Planetary Science* 40: 21–33.
- Kinslow R (ed.) (1970) *High Velocity Impact Phenomena*, 579pp. New York: Academic Press.
- Kopal Z (1966) The nature of secondary craters photographed by Ranger VII. *Icarus* 5: 201–213.
- Kuiper GP and Hartmann WK (1965) Interpretation of Ranger VII records: sec. C. *Jet Propulsion Laboratory, CalTech, Technical Reports 32-700*, pp. 59–63.
- La Feuvre M and Wieczorek MA (2006) The asymmetric cratering history of the terrestrial planets: Latitudinal effect. *Lunar and Planetary Science Conference 37th*: 1841.
- Le Feuvre M and Wieczorek MA (2005) The asymmetric cratering history of the moon. *Lunar and Planetary Science Conference 36th*: 2043.
- Levison HF, Dones L, Chapman CR, Stern SA, Duncan MJ, and Zahnle K (2001) Could the Lunar ‘Late Heavy Bombardment’ have been triggered by the formation of Uranus and Neptune? *Icarus* 151: 286–306.
- Lindblad BA, Neslusan L, Porubcan V, and Svoren J (2003) IAU meteor database of photographic orbits – version 2003. *Earth Moon and Planets* 93: 249–260.
- Lognonne P, Gagnepain-Beyneix J, and Chenet H (2003) A new seismic model of the Moon: Implications for structure, thermal evolution and formation of the Moon. *Earth and Planetary Science Letters* 211: 27–44.
- Love S and Ahrens TJ (1996) Catastrophic impacts on gravity dominated asteroids. *Icarus* 124: 141–155.
- Malin MC, Edgett KS, Posiolova LV, McColley SM, and Dobrea EZN (2006) Present-day impact cratering rate and contemporary gully activity on Mars. *Science* 314: 1573–1577.
- Marchi S, Morbidelli A, and Cremonese G (2005) Flux of meteoroid impacts on Mercury. *Astronomy and Astrophysics* 431: 1123–1127.
- McCrosky RE, Chao C-Y, and Posen A (1978) Prairie network fireball data. Part I: Summary of orbits. Center for Astrophysics Reprint No. 665. (Russian) *Meteoritika* 37: 44–59.
- McEwen AS and Bierhaus EB (2006) The importance of secondary cratering to age constraints on planetary surfaces. *Annual Review of Earth and Planetary Sciences* 34: 540–567.
- McEwen AS, Preble BS, Turtle EP, et al. (2005) The rayed crater Zunil and interpretations of small impact craters on Mars. *Icarus* 176: 351–381.
- McKinnon WB and Melosh HJ (1980) Evolution of planetary lithospheres: Evidence from multiringed basins on Ganymede and Callisto. *Icarus* 44: 454–471.
- McKinnon WB, Zahnle KJ, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Models and observations. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 969–1014. Tucson, AZ: University of Arizona Press.
- Melosh HJ (1989) *Impact Cratering – A Geologic Process*, 245 pp. New York: Oxford.
- Melosh HJ and Ivanov BA (1999) Impact crater collapse. *Annual Review of Earth and Planetary Sciences* 27: 385–415.
- Melosh HJ and McKinnon W (1978) The mechanics of ringed basin formation. *Geophysical Research Letters* 5: 985–988.
- Michel P, Froeschle C, and Farinella P (1996) Dynamical evolution of NEAs: Close encounters, secular perturbations and resonances. *Earth Moon and Planets* 72: 151–164.
- Michel P, Morbidelli A, and Bottke WF (2005) Origin and dynamics of near Earth objects. *Comptes Rendus Physique* 6: 291–301.
- Michel P and Thomas F (1996) The Kozai resonance for near-Earth asteroids with semimajor axes smaller than 2AU. *Astronomy and Astrophysics* 307: 310.
- Michel P and Yoshikawa M (2005) Earth impact probability of the Asteroid (25143) Itokawa to be sampled by the spacecraft Hayabusa. *Icarus* 179: 291–296.
- Milani A, Carpino M, Hahn G, and Nobili AM (1989) Dynamics of planet-crossing asteroids: Classes of orbital behavior: Project SPACEGUARD. *Icarus* 78: 212–269.
- Milani A and Farinella P (1995) An asteroid on the brink. *Icarus* 115: 209–212.
- Mizutani H, Kawakami S-I, Takagi Y, Kumazawa M, and Kato M (1983) Cratering experiments in sands and a trial for general scaling law. *Proceedings of the 13th Lunar and Planetary Science Conference, part 2, Journal of Geophysical Research, Supplement 88*: A835–A845.
- Moore HJ (1972) Ranger and other impact craters photographed by Apollo 16. *Apollo 16 Preliminary Science Report. NASA SP-315*, pp. 29_45–29_51. Washington, DC: NASA.
- Moore JM, Asphaug E, Belton MJS, et al. (2001) Impact features on Europa: Results of the Galileo Europa Mission (GEM). *Icarus* 151: 93–111.
- Morbidelli A (2002) *Modern Celestial Mechanics: Aspects of Solar System Dynamics*, 356 pp. New York: Taylor and Francis.
- Morbidelli A and Gladman B (1998) Orbital and temporal distributions of meteorites originating in the asteroid belt. *Meteoritics and Planetary Science* 33: 999–1016.
- Morbidelli A and Vokrouhlický D (2003) The Yarkovsky-driven origin of near-Earth asteroids. *Icarus* 163: 120–134.
- Morgan JV, Warner MR, Collins GS, Melosh HJ, and Christeson GL (2000) Peak-ring formation in large impact craters: Geophysical constraints from Chicxulub. *Earth and Planetary Science Letters* 183: 347–354.
- Morota T and Furumoto M (2003) Asymmetrical distribution of rayed craters on the Moon. *Earth and Planetary Science Letters* 206: 315–323.
- Morota T, Harujama J, and Furumoto M (2006) Lunar apex-antapex cratering asymmetry and origin of impactors in the Earth–Moon system. *Lunar and Planetary Science Conference 37th*: 1554.

- Morrison D, Harris AW, Sommer G, Chapman CR, and Carusi A (2002) Dealing with the impact hazard. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 739–754. Tucson, AZ: University of Arizona Press.
- Neukum G (1983) *Meteoritenbombardement und Datierung Planetarer Oberflächen. Habilitation Dissertation*, 186 pp. Munich: Ludwig-Maximilians-University.
- Neukum G and Hiller K (1981) Martian ages. *Journal of Geophysical Research* 86: 3097–3121.
- Neukum G and Ivanov BA (1994) Crater size distributions and impact probabilities on Earth from lunar, terrestrial-planet, and asteroid cratering data. In: Gehrels T (ed.) *Hazards due to Comets and Asteroids*, pp. 359–416. Tucson, AZ: University of Arizona Press.
- Neukum G, Koenig B, and Arkani-Hamed J (1975) A study of lunar impact crater size-distributions. *Moon* 12: 201–229.
- Neukum G, Ivanov BA, and Hartmann WK (2001a) Cratering records in the inner solar system in relation to the Lunar reference system. In: Kallenbach R, Griss J, and Hartmann WK (eds.) *Chronology and Evolution of Mars*, pp. 55–86. Dordrecht: Kluwer Academic Press.
- Neukum G, Oberst J, Hoffmann H, Wagner R, and Ivanov BA (2001b) Geologic evolution and cratering history of Mercury. *Planetary Space Science* 49: 1507–1521.
- Neukum G, Wagner R, Wolf U, et al. (1998) Bombardment history and ages of the Galilean satellites. *Annales Geophysicae* 16: 993.
- Neumann GA, Zuber MT, Wieczorek MA, McGovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research (Planets)* 109: 08002.
- Oberst J, Molau S, Heinlein D, et al. (1998) The ‘European Fireball Network’: Current status and future prospects. *Meteoritics Planetary Science* 33: 49–56.
- O’Brien DP and Greenberg R (2005) The collisional and dynamical evolution of the main-belt and NEA size distributions. *Icarus* 178: 179–212.
- O’Keefe JD and Ahrens TJ (1993) Planetary cratering mechanics. *Journal of Geophysical Research* 98: 17011–17028.
- O’Keefe JD and Ahrens TJ (1999) Complex craters: Relationship of stratigraphy and rings to impact conditions. *Journal of Geophysical Research* 104: 27091–27104.
- O’Keefe JD, Stewart ST, Lainhart ME, and Ahrens TJ (2001) Damage and rock-volatile mixture effects on impact crater formation. *International Journal of Impact Engineering* 26: 543–553.
- Onorato PIK, Uhlmann DR, and Simonds CH (1978) The thermal history of the Manicouagan impact melt sheet, Quebec. *Journal of Geophysical Research* 83: 2789–2798.
- Öpik EJ (1960) The frequency of crater diameters in Mare Imbrium. *Astronomical Journal* 65: 55.
- Öpik EJ (1976) *Interplanetary Encounters : Close-Range Gravitational Interactions*, 155 pp. Amsterdam: Elsevier.
- Pierazzo E and Melosh HJ (1999) Hydrocode modeling of Chicxulub as an oblique impact event. *Earth and Planetary Science Letters* 165: 163–176.
- Pierazzo E and Melosh HJ (2000) Understanding oblique impacts from experiments, observations, and modeling. *Annual Review of Earth and Planetary Sciences* 28: 141–167.
- Pierazzo E, Vickery AM, and Melosh HJ (1997) A reevaluation of impact melt production. *Icarus* 127: 408–423.
- Pieters CM, Tompkins S, Head JW, and Hess PC (1997) Mineralogy of the mafic anomaly in the South Pole-Aitken Basin: Implications for excavation of the lunar mantle. *Geophysical Research Letters* 24: 1903–1906.
- Pike RJ (1977a) Apparent depth/apparent diameter relation for lunar craters. In: Lunar Science Institute (ed.) *Proceedings of the 8th Lunar and Planetary Science Conference*, pp. 3427–3436. New York: Pergamon press.
- Pike RJ (1977b) Size-dependence in the shape of fresh impact craters on the moon. In: Roddy DJ, Pepin RO, and Merrill RB (eds.) *Impact and Explosion Cratering: Planetary and Terrestrial Implications*, pp. 489–509. New York: Pergamon Press.
- Pike RJ (1980) Control of crater morphology by gravity and target type – Mars, Earth, Moon. In: International Conference on Cold Fusion and Biberian J-P (eds.) *Proceedings of the 11th Lunar and Planetary Science Conference*, pp. 2159–2189. New York: Pergamon Press.
- Poag CW (1995) Late Eocene star wars: The Toms Canyon and Chesapeake Bay impact craters, US East Coast. *Meteoritics* 30: 562.
- Poag CW (1999) Secondary craters from the Chesapeake Bay impact. *Lunar and Planetary Science Conference*, 30th: 1047.
- Potts LV and von Frese RRB (2003a) Comprehensive mass modeling of the Moon from spectrally correlated free-air and terrain gravity data. *Journal of Geophysical Research (Planets)* 108d: 5-1–5-27.
- Potts LV and von Frese RRB (2003b) Crustal attributes of lunar basins from terrain-correlated free-air gravity anomalies. *Journal of Geophysical Research (Planets)* 108e: 3-1–3-21.
- Potts LV, von Frese RR, Leftwich TE, Taylor PT, Shum CK, and Li R (2004) Gravity-inferred crustal attributes of visible and buried impact basins on Mars. *Journal of Geophysical Research (Planets)* 109: 09009.
- Quinton C, Mangold N, Hartmann WK, and Allemand P (2007) Possible long-term decline in impact rates. Part 1: Martian geological data. *Icarus* 186: 1–10.
- Rabinowitz DL, Helin E, Lawrence K, and Pravdo S (2000) A reduced estimate of the number of kilometer-sized Near-Earth Asteroids. *Nature* 403: 165–166.
- Richardson JE, Melosh HJ, Greenberg RJ, and O’Brien DP (2005) The global effects of impact-induced seismic activity on fractured asteroid surface morphology. *Icarus* 179: 325–349.
- Roberts WA (1964) Secondary craters. *Icarus* 3: 348–364.
- Roddy DJ, Pepin RO, and Merrill RB (eds.) (1977) *Impact and Explosion Cratering*, 1301 pp. New York: Pergamon Press.
- Rubincam DP (1998) Yarkovsky thermal drag on small asteroids and Mars–Earth delivery. *Journal of Geophysical Research* 103: 1725.
- Rubincam DP (2000) Radiative spin-up and spin-down of small asteroids. *Icarus* 148: 2–11.
- Ruskol EL (1966) On the past history of the Earth–Moon system. *Icarus* 5: 221–227.
- Ryan EV and Melosh HJ (1998) Impact fragmentation: From the laboratory to asteroids. *Icarus* 133: 1–24.
- Safronov VS (1969) *Evolutsia Doplanetnogo Oblaka I Obrazovanie Zemli i Planet (Evolution of the Protoplanetary Cloud and Formation of the Earth and Planets)*. Moscow: Nauka Press, 244 pp. (English translation (1972) NASA Technical Translation F-677, US Department of Commerce, Springfield, VA 22151 and Israel Program for Scientific Translation Jerusalem, 206 pp.).
- Schenk PM (2002) Thickness constraints on the icy shells of the galilean satellites from a comparison of crater shapes. *Nature* 417: 419–421.
- Schenk PM, Chapman CR, Zahnle K, and Moore JM (2004) Ages and interiors: The cratering record of the Galilean satellites. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter The Planet, Satellites and Magnetosphere*, pp. 427–456. Cambridge, UK: Cambridge University Press.
- Schenk PM and Ridolfi FJ (2002) Morphology and scaling of ejecta deposits on icy satellites. *Geophysical Research Letters* 29: 31-1–31-4.
- Schmidt RM (1980) Meteor crater: Energy of formation – Implications of centrifuge scaling. In: International

- Conference on Cold Fusion and Biberian J-P (eds.) *Proceedings of the 11th Lunar and Planetary Science Conference*, pp. 2099–2128. New York: Pergamon Press.
- Schmidt RM (1987) Preliminary scaling results for crater rim-crest diameter. *Lunar and Planetary Science Conference 18th*: 878–879.
- Schmidt RM and Housen KR (1987) Some recent advances in the scaling of impact and explosion cratering. *International Journal of Impact Engineering* 5: 543–560.
- Settle M (1980) The role of fallback ejecta in the modification of impact craters. *Icarus* 42: 1–19.
- Sheppard SS (2006) Solar system: A planet more, a planet less? *Nature* 439: 541–542.
- Shoemaker EM (1962) Interpretation of lunar craters. In: Kopal Z (ed.) *Physics and Astronomy of the Moon*, pp. 283–359. New York: Academic Press.
- Shoemaker EM (1963) Impact mechanics at Meteor Crater, Arizona. In: Middlehurst BM and Kuiper GP (eds.) *The Moon, Meteorites and Comets*, pp. 301–336. Chicago, Ill: University of Chicago Press.
- Shoemaker EM (1965) Preliminary analysis of the fine structure of the lunar surface in Mare Cognitum. *Jet Propulsion Laboratory, CalTech, Technical Report 32-700*, pp. 75–134.
- Shoemaker EM (1977) Astronomically observable crater-forming projectiles. In: Roddy DJ, Pepin RO, and Merrill RB (eds.) *Impact and Explosion Cratering: Planetary and Terrestrial Implications*, pp. 617–628. New York: Pergamon Press.
- Shoemaker EM and Wolfe RF (1982) Cratering time scales for the Galilean satellites. In: Morrison D (ed.) *Satellites of Jupiter*, pp. 277–339. Tucson, AZ: University of Arizona Press.
- Shoemaker EM and Shoemaker CS (1990) Proterozoic impact record of Australia. *Abstracts for the International Workshop on Meteorite Impact on the Early Earth. LPI Contributions 746*, pp. 47–48. Houston, TX: Lunar and Planetary Institute.
- Shoemaker EM, Wolfe RF, and Shoemaker CS (1990) Asteroid and comet flux in the neighborhood of Earth. In: Sharpton VL and Ward PD (eds.) *Global Catastrophes in Earth History, GSA SP-247*, pp. 155–170. Boulder, CO: Geological Society of America.
- Shoemaker EM, Weissman PR, and Shoemaker CS (1994) The flux of periodic comets near Earth. In: Gehrels T (ed.) *Hazards Due to Comets and Asteroids*, pp. 313–335. Tucson, AZ: University of Arizona Press.
- Shoemaker EM, Hait MH, Swann GA, et al. (1970) Lunar regolith at Tranquillity Base. *Science* 167: 452.
- Spudis PD (1993) *The Geology of Multi-Ring Impact Basins*, 263 pp. Cambridge: Cambridge University Press.
- Stöffler D and Ryder G (2001) Stratigraphy and isotope ages of lunar geologic units: Chronological standard for the inner solar system. *Space Science Reviews* 96: 9–54.
- Stöffler D, Artemieva NA, Ivanov BA, et al. (2004) Origin and emplacement of the impact formations at Chicxulub, Mexico, as revealed by the ICDP deep drilling at Yaxcopoil-1 and by numerical modeling. *Meteoritics Planetary Science* 39: 1035–1067.
- Strom RG, Croft SK, and Barlow NG (1992) The Martian impact cratering record. In: Kieffer HH, Jakovsky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 383–423. Tucson, AZ: University of Arizona Press.
- Strom RG and Neukum G (1988) The cratering record on Mercury and the origin of impacting objects. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 336–373. Tucson, AZ: University of Arizona Press.
- Stuart JS (2001) A near-Earth asteroid population estimate from the LINEAR survey. *Science* 294: 1691–1693.
- Stuart JS and Binzel RP (2004) Bias-corrected population, size distribution, and impact hazard for the near-Earth objects. *Icarus* 170: 295–311.
- Sugano T and Heki K (2004) Isostasy of the Moon from high-resolution gravity and topography data: Implication for its thermal history. *Geophysical Research Letters* 31: 24703.
- Toksoz MN, Press F, Anderson K, et al. (1972) Velocity structure and properties of the lunar crust. *Moon* 4: 490.
- Touma J (2000) The phase space adventure of the Earth and Moon. In: Canup RM and Righter K (eds.) *Origin of the Earth and Moon*, pp. 165–178. Tucson, AZ: University of Arizona Press.
- Tsiganis K, Gomes R, Morbidelli A, and Levison HF (2005) Origin of the orbital architecture of the giant planets of the Solar System. *Science* 305: 459–461.
- Ugalde HA, Artemieva N, and Milkereit B (2005) Magnetization on impact structures – Constraints from numerical modeling and petrophysics. In: Kenkemann T, Horz F, and Deutsch A (eds.) *Special paper 384: Large Meteorite Impacts III*, pp. 25–42. Boulder, CO: Geological Society of America.
- Vanorio T, Prasad M, Patella D, and Nur A (2002) Ultrasonic velocity measurements in volcanic rocks: Correlation with microtexture. *Geophysical Journal International* 149: 22–36.
- Vokrouhlický D, Broz M, Bottke WF, Nesvorný D, and Morbidelli A (2006a) Yarkovsky/YORP chronology of asteroid families. *Icarus* 182: 118–142.
- Vokrouhlický D, Broz M, Morbidelli A, Bottke WF, Nesvorný D, et al. (2006b) Yarkovsky footprints in the Eos family. *Icarus* 182: 92–117.
- Ward WR (1992) Long-term orbital and spin dynamics of Mars. In: Kieffer HH, Jakovsky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 298–320. Tucson, AZ: University of Arizona Press.
- Warren PH (2001) Porosities of lunar meteorites: Strength, porosity, and petrologic screening during the meteorite delivery process. *Journal of Geophysical Research* 106: 10101–10112.
- Weibull W (1951) A statistical distribution function of wide applicability. *Journal of Applied Mechanics* 18: 293–297.
- Weissman PR, Bottke WF, and Levison HF (2002) Evolution of comets into asteroids. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 669–686. Tucson, AZ: University of Arizona Press.
- Werner SC, Harris AW, Neukum G, and Ivanov BA (2002) NOTE: The near-Earth asteroid size-frequency distribution: A snapshot of the lunar impactor size-frequency distribution. *Icarus* 156: 287–290.
- Werner SC, Ivanov BA, and Neukum G (2006) Secondary cratering and age determination on Mars. *LPSC 37th* 37: #1595.
- Werner SC, van Gasselt S, and Neukum G (2003) Continual geological activity in Athabasca Valles, Mars. *Journal of Geophysical Research (Planets)* 108: 8081.
- Wetherill GW (1967) Collisions in the asteroid belt. *Journal of Geophysical Research* 72: 2429–2444.
- Wetherill GW (1978) Accumulation of the terrestrial planets. In: Gehrels T (ed.) *Protostars and Planets: Studies of Star Formation and of the Origin of the Solar System*, pp. 565–598. Tucson, AZ: University of Arizona Press.
- Wetherill GW (1980) Formation of the terrestrial planets. *Annual Review of Astronomy and Astrophysics* 18: 77–113.
- Wieczorek MA and Phillips RJ (1999) Lunar multiring basins and the cratering process. *Icarus* 139: 246–259.
- Witaker EA (1972) Artificial lunar impact craters: Four new identifications. *Apollo 16 Preliminary Science Report. NASA SP-315*, pp. 29_39–29_45. Washington, DC: NASA.
- Working_Group (including Avidson RE, Boyce J, Chapman C, et al.) (1979) Standard techniques for presentation and analysis of crater size-frequency data. *Icarus* 37: 467–474.

- Wünnemann K, Collins GS, and Melosh HJ (2006) A strain-based porosity model for use in hydrocode simulations of impacts and implications for transient crater growth in porous targets. *Icarus* 180: 514–527.
- Yakovlev OI and Basilevsky AT (1994) Experimental studies of geochemical aspects of impact cratering. In: Dressler B, Grieve RAF, and Sharpton VL (eds.) *Geological Society of America Special Paper 293: Large Meteorite Impacts and Planetary Evolution*, pp. 73–80. Boulder, CO: GSA.
- Zahnle K, Dones L, and Levison HF (1998) Cratering rates on the Galilean satellites. *Icarus* 136: 202–222.
- Zahnle K, Schenk P, Levison H, and Dones L (2003) Cratering rates in the outer Solar System. *Icarus* 163: 263–289.
- Zeldovich IB (1956) Gas motion due to a short-time pressure impulse (impact). *Akusticheskii Zhurnal* 21: 28–38.
- Zharkov VN and Kozenko AV (1990) On the role of Jupiter in the formation of the giant planets. *Soviet Astronomy Letters* 16: 73.
- Zieg MJ and Marsh BD (2005) The Sudbury igneous complex: Viscous emulsion differentiation of a superheated impact melt sheet. *GSA Bulletin* 117: 1427–1450.
- Zuber MT, Solomon SC, et al. (2000) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

Relevant Websites

<http://cfa-www.harvard.edu> – Harvard Smithsonian Center for Astrophysics

<http://www.muses-c.isas.ac.jp> – Recent Status of Hayabusa Spacecraft

10.07 Planetary Magnetism

J. E. P. Connerney, NASA Goddard Space Flight Center, Greenbelt, MD, USA

Published by Elsevier B.V.

10.07.1	Introduction	243
10.07.2	Tools	245
10.07.2.1	The Offset Tilted Dipole	245
10.07.2.2	Spherical Harmonic Models	246
10.07.3	Terrestrial Planets	248
10.07.3.1	Mercury	248
10.07.3.1.1	Observations	248
10.07.3.1.2	Models	249
10.07.3.2	Venus	250
10.07.3.2.1	Observations	251
10.07.3.2.2	Discussion	251
10.07.3.3	Mars	251
10.07.3.3.1	Observations	251
10.07.3.3.2	Models	254
10.07.3.3.3	Discussion	256
10.07.4	Gas Giants	256
10.07.4.1	Jupiter	256
10.07.4.2	Observations	257
10.07.4.3	Models	258
10.07.5	Discussion	262
10.07.5.1	Saturn	263
10.07.5.1.1	Observations	263
10.07.5.1.2	Models	264
10.07.5.1.3	Discussion	265
10.07.6	Ice Giants	267
10.07.6.1	Uranus	267
10.07.6.1.1	Models	267
10.07.6.2	Neptune	269
10.07.7	Satellites and Small Bodies	272
10.07.7.1	Moon	272
10.07.7.2	Ganymede	273
10.07.8	Discussion	274
10.07.9	Summary	275
	References	275

10.07.1 Introduction

Planetary magnetism, beginning with the study of the geomagnetic field, is one of the oldest scientific disciplines. Magnetism was both a curiosity and a useful tool for navigation, the subject of myth and superstition as much as scientific study. With the publication of *De Magnete* in 1600, William Gilbert demonstrated that the Earth's magnetic field was like that of a bar magnet. Gilbert's treatise on magnetism established the global

nature and dipole geometry of the field and led to an association between magnetism and the interior of planet Earth. Observations of Earth's magnetic field gathered over time demonstrated that it varies slowly with time. The orientation of the field was observed to execute a long-term or secular drift called polar wander. Later, when accurate measurements of the magnitude of the field could be made, secular variations in the field magnitude were identified. Given the importance of accurate magnetic field maps in

navigation, global surveys of the magnetic field were initiated periodically, at first by ships of discovery and commerce and later by specialized survey ships built to minimize distortion of the magnetic field they sought to measure. In recent times, polar-orbiting, near-Earth satellites such as MAGSAT ([Langel et al., 1982](#)), Orsted, CHAMP, and SAC-C provided the capability to map the magnetic field over the entire surface of Earth in a brief period, to capture a snapshot of the field at each epoch. By studying the magnetic properties of terrestrial rocks, paleomagnetists can extend the history of the geodynamo over timescales of hundreds of millions of years. For a discussion of the magnetism of the Earth see volume 5 of this treatise, for a discussion of the geodynamo *see* Chapters 8.03, 8.08, and 8.09, for a discussion of planetary dynamos *see* Chapter 10.08.

Planetary magnetism is also a recent study, born of the space age and a new era of discovery that began in the 1960s with the launch of spacecraft to other bodies. The Sun and stars were known to possess magnetic fields, but one could only speculate on whether the planets and satellites in our solar system were magnetized. The planet Jupiter is an exception; its magnetic field was inferred from the detection of nonthermal radio emissions ([Burke and Franklin, 1955](#)) as early as 1955, before the new age of exploration. In the decades that followed, the gross characteristics of Jupiter's magnetic field were deduced from variations of the plane of polarization of the emissions or variations in flux density (see, e.g., the review by [Berge and Gulkis \(1976\)](#)). Estimates of the orientation of Jupiter's field (dipole tilt) were more successful than those of its magnitude; the detailed nature of Jupiter's field was established by direct measurement, by spacecraft.

In little more than three decades, spacecraft have visited nearly every large body in the solar system. The first *in situ* observations of the Jovian magnetosphere were obtained by the Pioneer 10 and 11 spacecraft in December 1973 and December 1974, respectively ([Smith et al., 1974, 1975](#); [Acuña and Ness, 1976](#)). The Pioneer 11 spacecraft continued to make the first observations of Saturn's magnetic field as well ([Smith et al., 1980a, 1980b](#); [Acuña and Ness, 1980](#)). Voyagers 1 and 2 followed, passing through the Jovian magnetosphere in March and July of 1979, respectively ([Ness et al., 1979a, 1979b](#)). Voyagers 1 and 2 continued onward to Saturn, obtaining additional observations of Saturn's magnetic field in November 1980 and August 1981 ([Ness et al., 1981, 1982](#)). Voyager 2 continued on its 'grand tour' of the solar system, with the discovery of the magnetic field of Uranus in January 1986 ([Ness et al., 1986](#)) and that of Neptune in August 1989 ([Ness et al., 1989](#)).

Voyager 2's remarkable tour of the outer solar system was made possible by the celestial alignment of the planets, the syzygy, which occurs every 187 years or so.

Jupiter was visited again, in February 1992, by the Ulysses spacecraft *en route* to the high-latitude heliosphere and again (albeit at great distance) in 2004. Ulysses used a Jupiter gravity assist to escape the ecliptic plane, but in so doing passed to within $\sim 6 R_J$ ([Balogh et al., 1992](#)). Additional observations were obtained by the Galileo Orbiter beginning in 1995 ([Johnson et al., 1992](#)), but apart from the de-orbit maneuver to send Galileo into Jupiter's atmosphere for planetary protection (of the Galilean satellites, not Jupiter), little additional information regarding Jupiter's main field was obtained. Instead, Galileo surprised us with the remarkable discovery of an intrinsic magnetic field of Ganymede ([Kivelson et al., 1996](#)) and the induced magnetic fields of Europa and Callisto ([Khurana et al., 1998](#); [Zimmer et al., 2000](#)). Today, the Cassini Orbiter orbits Saturn gathering observations of the distant magnetosphere; close in observations must await the polar orbits scheduled for late in the mission.

The Mariner 10 spacecraft discovered Mercury's magnetic field in brief flybys in 1974 and 1975 ([Ness et al., 1974a](#); [Ness et al., 1975, 1976](#)) and provided the first indication that slowly rotating Venus was not magnetized ([Ness et al., 1974b](#)). A great many spacecraft visited Mars ([Connerney et al., 2005](#)), beginning with the Soviet launch of Mars 1 in 1962, but Mars did not relinquish his secrets until the Mars Global Surveyor (MGS) spacecraft entered orbit about Mars in September of 1997. MGS measurements demonstrated that Mars does not have a global magnetic field of internal origin, but its crust is intensely magnetized ([Acuña et al., 1998](#)). So Mars once had an internal magnetic field, for at least as long as it took the crust to acquire remanence. Perhaps one-day paleomagnetic measurements will reveal something of the history of that dynamo. The only traditional planet (setting aside the debate on the definition of 'planet') not yet visited by spacecraft is the Pluto–Charon system, which is scheduled for a visit by the New Horizons spacecraft in July 2015. The New Horizons spacecraft is not instrumented to measure magnetic fields, so no direct measurement is possible, but the presence of a magnetic field may be inferred from measurements made by other instruments.

Our knowledge of the magnetic fields of the planets is, in large part, based upon the planetary flybys

that occurred decades ago. Mars is the exception, having been mapped thoroughly by the polar-orbiting MGS spacecraft at a nominal altitude of 400 km; elsewhere the observations are sparse in both space and time, particularly considered in light of the data available for the study of the geomagnetic field. They are, however, sufficient to describe a diverse group of planetary magnetic fields and provide clues to guide the continued development of dynamo theory.

10.07.2 Tools

A planet with an intrinsic magnetic field stands as an obstacle to the solar wind, the high-velocity stream of plasma emanating from the sun. The interaction of the planetary field and the solar wind forms a multilayered interaction region, often approximated as a set of conic sections. The supersonic solar wind forms a shock upstream of the obstacle ('bow shock'). The slowed solar wind flows around the obstacle within the 'magnetosheath', bounded by the bow shock and the 'magnetopause', often approximated by a paraboloid of revolution about the planet–Sun line. Within the magnetopause is a region dominated by the planetary field, called the 'magnetosphere', a term introduced by Tom Gold in the post-war era of rocketry that provided access to space in the early 1950s. The magnetospheric magnetic field extends well downstream in the anti-sunward direction, as if stretched like an archer's bowstring, to form the 'magnetotail'. The observed magnetic field can be regarded as the sum of contributions from several sources, dominated by the planetary dynamo, sometimes referred to as the 'main field'. Other relatively minor sources include currents flowing on the magnetopause and tail currents arising from the solar wind interaction and distributed currents ('ring currents') due to the motion of charged particles within the magnetosphere. Field-aligned currents, called 'Birkeland currents', flow between the magnetosphere and the planet's electrically conducting ionosphere, particularly during magnetospheric storms, leading to intense auroral displays. All of these currents produce magnetic fields, often relatively small when compared to the planetary field, particularly close to the planet's surface. However, if an accurate model is to be obtained from limited or distant observations, these sources can be a significant source of errors if not modeled.

The planetary field is most often characterized by one of a very few simple models, depending on the

application and/or the availability of observations. Simple representations include dipole models of one form or another and a potential field representation as a series of spherical harmonics. External fields, or those arising from currents outside the planet, are often accommodated by explicit models or spherical harmonic methods. The latter approach is useful only in regions free of currents.

10.07.2.1 The Offset Tilted Dipole

Dipole models have found use in the interpretation of magnetic field observations, dating to Gilbert's analysis of Earth's magnetic field. The dipole is the simplest approximation to a localized source. The vector magnetic field $\mathbf{B}(\mathbf{r})$ of a dipole at the origin is given by

$$\mathbf{B}(\mathbf{r}) = \left[-\frac{\mathbf{m}}{r^3} + \frac{3(\mathbf{m} \cdot \mathbf{r})\mathbf{r}}{r^5} \right]$$

where \mathbf{m} is the magnetic moment in units of $B - r^3$ and $r = |\mathbf{r}|$. The field of a dipole located near the origin is computed by translation of the origin of the coordinate system. Offset tilted dipole (OTD) models are found using forward modeling techniques to compare *in situ* observations to the OTD model field. The dipole offset (x_0 , y_0 , z_0) is not linearly related to the computed field, so an iterative procedure is often used to find the optimal dipole position (e.g., Smith *et al.*, 1976). Alternatively one may perform a spherical harmonic analysis (see next section) of degree and order 1 centered upon each of a large number of candidate dipole positions, selecting the offset which minimizes the model residuals. In either case, the result is the OTD which most closely approximates the field measured along the spacecraft trajectory.

The simple OTD representation provides a convenient approximation to the field, particularly useful at larger distances. It is useful in visualizing the gross characteristics of the field and it is often used to chart the motion of charged particles throughout the magnetosphere. However, the simplicity of the OTD is also a limitation, in that complex field geometries cannot be adequately represented with such a limited parametrization. In general, a dipole representation can be expected to closely approximate the field for radial distances $r \gg a$, where a is a characteristic source dimension. At lesser radial distances, an accurate description of the field requires more flexibility in parametrization, for

example, consideration of higher-degree and -order moments.

A logical extension of the simple OTD model is one of many dipoles. Earth's field has been approximated by adding one or more smaller dipoles to a 'main dipole' and treating the field as a summation over distinct sources; this provides a simple means of introducing an anomaly, or departure from dipole geometry. The Earth's field has been approximated by an ensemble of many dipoles located on the surface of a sphere of radius $0.5 R_e$, the fluid core radius. In addition to the field generated deep within Earth, induced and remanent magnetization in the crust can be modeled with another set of dipoles, scattered about a sphere of radius $1.0 R_e$ to approximate sources in the crust (Mayhew and Estes, 1983). A similar approach has been taken to model the crustal magnetic field of Mars (Purucker *et al.*, 2000).

10.07.2.2 Spherical Harmonic Models

In the absence of local currents ($\nabla \times \mathbf{B} = 0$), the magnetic field may be expressed as the gradient of a scalar potential $V(\mathbf{B} = -\nabla V)$. It is particularly advantageous to expand the potential V in a series of functions, called spherical harmonics, which are solutions to Laplace's equation in spherical coordinates. This approach, introduced by Gauss in 1839, has been very popular in studies of the geomagnetic field, and subsequently those of the planets. The traditional spherical harmonic expansion of V is given by (e.g., Chapman and Bartels, 1940; Langel, 1987)

$$V = a \sum_{n=1}^{\infty} \left\{ \left(\frac{r}{a}\right)^n T_n^e + \left(\frac{a}{r}\right)^{n+1} T_n^i \right\}$$

where a is the planet's equatorial radius. The first series in increasing powers of r represents contributions due to external sources, with

$$T_n^e = \sum_{m=0}^n \left\{ P_n^m(\cos \theta) [G_n^m \cos(m\phi) + H_n^m \sin(m\phi)] \right\}$$

The second series in inverse powers of r represents contributions due to the planetary field or internal sources, with

$$T_n^i = \sum_{m=0}^n \left\{ P_n^m(\cos \theta) [g_n^m \cos(m\phi) + b_n^m \sin(m\phi)] \right\}$$

The $P_n^m(\cos \theta)$ are Schmidt quasi-normalized associated Legendre functions of degree n and order m , and the g_n^m , b_n^m and G_n^m , H_n^m are the internal and

external Schmidt coefficients, respectively. These are most often presented in units of Gauss or nanoteslas ($1 \text{ G} = 10^5 \text{ nT}$) for a particular choice of equatorial radius a of the planet. Different values of the equatorial radius have been used and should be noted in the comparison of various field models. The angles θ and ϕ are the polar angles of a spherical coordinate system, θ (co-latitude) measured from the axis of rotation and ϕ increasing in the direction of rotation. The three components of the magnetic field (internal field only) are obtained from the expression for V above:

$$B_r = -\frac{\partial V}{\partial r} = -\sum_{n=1}^{\infty} \sum_{m=0}^n \left\{ (n+1) \left(\frac{a}{r}\right)^{n+2} [g_n^m \cos(m\phi) + b_n^m \sin(m\phi)] P_n^m(\cos \theta) \right\}$$

$$B_\theta = -\frac{\partial V}{r \partial \theta} = -\sum_{n=1}^{\infty} \sum_{m=0}^n \left\{ \left(\frac{a}{r}\right)^{n+2} [g_n^m \cos(m\phi) + b_n^m \sin(m\phi)] \frac{dP_n^m(\cos \theta)}{d\theta} \right\}$$

$$B_\phi = -\frac{-1}{r \sin \theta} \frac{\partial V}{\partial \theta} = \frac{1}{\sin \theta} \sum_{n=1}^{\infty} \sum_{m=0}^n \left\{ m \left(\frac{a}{r}\right)^{n+2} [g_n^m \cos(m\phi) - b_n^m \sin(m\phi)] P_n^m(\cos \theta) \right\}$$

The leading terms in the series (through degree 4) may be computed using the Schmidt quasi-normalized Legendre functions listed in **Table 1**.

Observations are often rendered in a west longitude system, in which the longitude of a stationary observer increases in time as the planet rotates. West longitudes are simply related to the angle ϕ by $\lambda = 360 - \phi$. Longitudes for the terrestrial planets may be assigned by reference to surface features. For the outer planets, the rotation period is inferred from observations of episodic radio emissions, which, by assumption, are locked in phase with the rotation of the planet's magnetic field and hence the electrically conducting deep interior of the planet. For these planets, longitudes must be assigned with knowledge of the rotation rate and time. For Jupiter, observation of high-frequency radio emission (penetrating Earth's ionosphere) over several decades led to an accurate determination of the planet's rotation period, although one must be cognizant of the occasional update (e.g., $\lambda_{\text{III}}(1957)$ vs. $\lambda_{\text{III}}(1965)$). These two systems result in significantly different longitudes when observations of the Pioneer era are compared with those of the Voyager era, so one must

Table 1 Schmidt quasi-normalized Legendre functions**Legendre functions**

$P_1^0 = (1) \cos \theta$
$P_1^1 = (1) \sin \theta$
$P_2^0 = \left(\frac{3}{2}\right) (\cos^2 \theta - 1/3)$
$P_2^1 = (3^{1/2}) \cos \theta \sin \theta$
$P_2^2 = \left(\frac{3^{1/2}}{2}\right) \sin^2 \theta$
$P_3^0 = \left(\frac{5}{2}\right) \cos \theta (\cos^2 \theta - 9/15)$
$P_3^1 = \left(\frac{5(3)^{1/2}}{2(2)^{1/2}}\right) \sin \theta (\cos^2 \theta - 3/15)$
$P_3^2 = \left(\frac{(3)^{1/2}(5)^{1/2}}{2}\right) \cos \theta \sin^2 \theta$
$P_3^3 = \left(\frac{(5)^{1/2}}{2(2)^{1/2}}\right) \sin^3 \theta$
$P_4^0 = \left(\frac{35}{8}\right) [\cos^4 \theta - (30/35) \cos^2 \theta + (3/35)]$
$P_4^1 = \left(\frac{7(5)^{1/2}}{2(2)^{1/2}}\right) \cos \theta \sin \theta (\cos^2 \theta - 3/7)$
$P_4^2 = \left(\frac{7(5)^{1/2}}{4}\right) \sin^2 \theta (\cos^2 \theta - 1/7)$
$P_4^3 = \left(\frac{(7)^{1/2}(5)^{1/2}}{2(2)^{1/2}}\right) \cos \theta \sin^3 \theta$
$P_4^4 = \left(\frac{(7)^{1/2}(5)^{1/2}}{8}\right) \sin^4 \theta$

The leading term (in parentheses) converts the functions from Gaussian normalization, convenient for recursive computations, to Schmidt quasi-normalized form.

be careful to transform earlier results to a current coordinate system. A detailed description of Jovian coordinate systems is given by [Dessler \(1983\)](#). For Saturn, Uranus, and Neptune the rotation periods and longitude system in current use are derived from Voyager observations of radio emissions ([Desch and Kaiser, 1981](#); [Desch et al., 1986](#); [Warwick et al., 1989](#)). In practice (e.g., Saturn) this can lead to significant ambiguity in longitudes and great confusion in reconciliation of observations taken over a lengthy span of time.

The expansion in increasing powers of r , representing external fields, is often truncated at $n = N_{\max} = 1$ corresponding to a uniform external field due to distant magnetopause and tail currents, that is, sources well beyond the region of interest. No potential field can represent the field of local

currents, so in practice the external expansion is often truncated at $N_{\max} = 1$ and local currents are modeled with the aid of explicit models. The dominant contribution from local currents may be described as a large-scale equatorial current system ([Connerney et al., 1981](#)). Field-aligned, or Birkeland currents, may also play a role in addition to induction effects, but thus far have not been included in such analyses, excepting those of Earth (e.g., [Backus, 1986](#); [Sabaka et al., 2002](#)). It is very important to adequately treat the fields of local currents, because even small errors accumulating along the spacecraft trajectory can lead to rather large errors in the spherical harmonic coefficients deduced from observations.

The maximum degree and order required of the internal field expansion depends on the complexity of the field measured. In usual practice, the series above is truncated at a maximum degree N_{\max} , where N_{\max} is large enough to follow variations in the field at the orbital altitude of the measurement. The number of free parameters grows rapidly with increasing N_{\max} , as $n_p = (N_{\max} + 1)^2 - 1$. If the observations are well distributed on a sphere, the spherical harmonics do not covary, and the coefficients obtained are independent of the truncation of the series. However, if the observations are poorly distributed or sparse, as is often the case for a planetary flyby, the usual spherical harmonics covary, so it is advisable to construct new orthogonal basis functions. Furthermore, if an arbitrarily small choice of N_{\max} is imposed upon the model in order to obtain a manageable linear system, large errors in low-degree and -order terms will result from the neglect of higher-order terms. To address these problems, [Connerney \(1981\)](#) introduced a method of analysis based on the singular-value decomposition (SVD) of Lanczos. The method involves the construction of partial solutions to the generalized linear inverse problem to obtain estimates of those parameters which are constrained by the observations. The model parameters which are not constrained by the data are readily identified and exploited in characterization of model nonuniqueness. One advantage of this approach is that the physical model of the planetary field (expansion to degree N_{\max}) does not depend on the completeness or extent of the available observations. Partial solutions, and estimates of the model parameters that result, may be interpreted within the context of the ‘resolution matrix’ for the particular solution (see [Connerney, 1981](#); [Connerney et al., 1991](#)). The close flyby of Neptune in 1989 serves to illustrate this concept well.

10.07.3 Terrestrial Planets

10.07.3.1 Mercury

Mercury, the smallest ($R_m = 2439$ km) and least massive ($M_m = 0.056M_e$) planet in the solar system, occupies the pole position in its race about the Sun, with a semimajor axis of 0.39 AU. Due to its large orbital eccentricity (0.206), in one orbit about the Sun (88 days) its orbital radius varies from 0.31 to 0.47 AU. It is locked in a spin-orbit resonance (Pettengill and Dyce, 1965; Goldreich and Peale, 1966) rotating once every 58.65 days or three times every two orbital periods. It is difficult to observe from Earth, due to its proximity to the Sun, and little was known about Mercury prior to observation by spacecraft on 29 March 1974, the first of three encounters with the Mariner 10 spacecraft. Discovery of the intrinsic magnetic field of Mercury was a bit surprising since it was widely assumed that a body as small as Mercury would have completely solidified, ruling out generation of a magnetic field by dynamo action (but see Chapter 10.09). Here we describe only the Mariner 10 observations of Mercury, the only *in situ* observations obtained in the first four decades of space exploration. As this chapter is being written, the Messenger spacecraft is on its way to Mercury (Solomon *et al.*, 2001; Gold *et al.*, 2001), in part to better understand the internal magnetic field and its origin. It will enter orbit in March 2011, after a series of flybys of Earth, Venus (two flybys), and Mercury (three flybys: January and October 2008, and September 2009). This mission may be expected to rewrite the book on Mercury and its magnetic field. A more ambitious mission to Mercury, ‘BepiColombo’, is in development by the European Space Agency (ESA) and the Institute of Space and Astronautical Science in Japan (ISAS) for launch in the near future (Anselmi and Scoon, 2001).

10.07.3.1.1 Observations

Mariner 10, known pre-launch as the Mariner Venus–Mercury spacecraft, was launched on 3 November 1973. The spacecraft used a close flyby of Venus on 5 February 1974 (for gravity assist) to achieve three close encounters with Mercury (29 March 1974, 21 September 1974, and 16 March 1975) before it exhausted its expendables. The orbital period of Mariner 10 after the Venus gravity assist was 176 days, very nearly commensurate with Mercury’s orbital (and rotation) period, providing three very similar close encounters. Each time the

spacecraft approached the planet from a local time of approximately 1900 h and departed near 0700 h. Given the brevity of these encounters, the relatively slow planetary rotation rate, and the orbital commensurabilities, each encounter approximates a shot past a stationary object, the major difference being the altitude and subspacecraft latitude of the point of closest approach. The first and third of these encounters provided observations within the Hermean magnetosphere (Figure 1); the second was targeted well above the sunlit hemisphere and beyond the reach of the Hermean magnetosphere in the upstream direction of the solar wind.

Mariner 10’s first flyby was relatively close, as planetary flybys go, with a periapsis of $1.29 R_m$ over the darkened hemisphere (anti-sunward direction), passing from dusk to dawn behind the planet as viewed from the Sun. Mariner 10 traveled from bow shock (inbound) to bow shock (outbound) in a mere 33 min, 17 of which were spent within the confines of a Hermean magnetosphere (Figure 2). These observations are best understood in terms of the field external to the planet, even at such a close distance. The field increased in magnitude from about 45 nT just inside the magnetopause to approximately 100 nT, all the while maintaining the anti-sunward

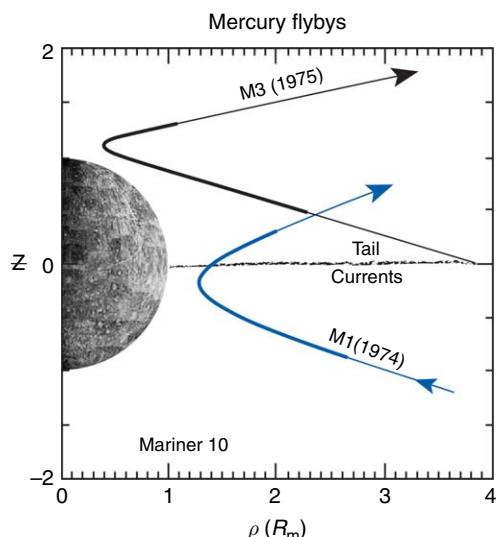


Figure 1 Mariner 10 trajectory in Mercury-centered cylindrical coordinates for the first (M1) and third (M3) encounters. The spacecraft distance from the equator (z) as a function of distance from the z -axis (ρ) is given in units of Mercury radius. The part of the trajectory within the Hermean magnetosphere is indicated by the thick line segment.

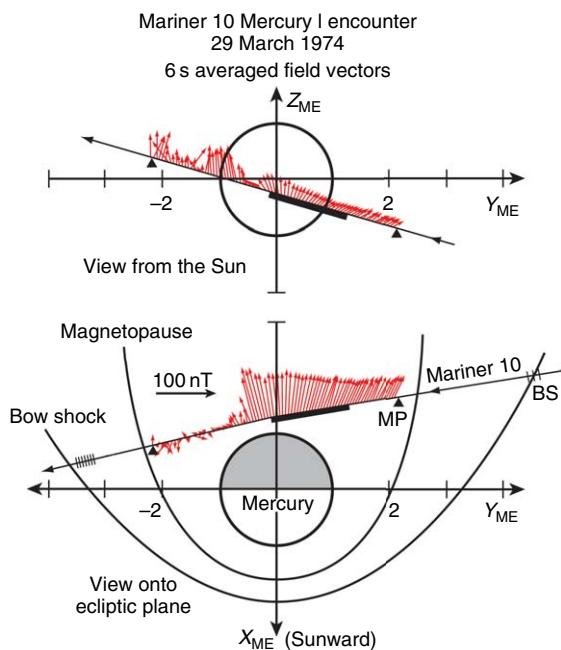


Figure 2 M1 encounter trajectory and magnetic field observations in Mercury ecliptic coordinates, viewed from the Sun (top) and viewed along Mercury's rotation axis (bottom). The vector magnetic field observed along the Mariner 10 trajectory is illustrated by vector projection onto the z-y (top) and x-y (bottom) planes; vectors originate at the spacecraft position in ME coordinates at the time of observation. Observed bow shock (BS) and magnetopause (MP) positions are shown along with scaled nominal bow shock and magnetopause boundaries.

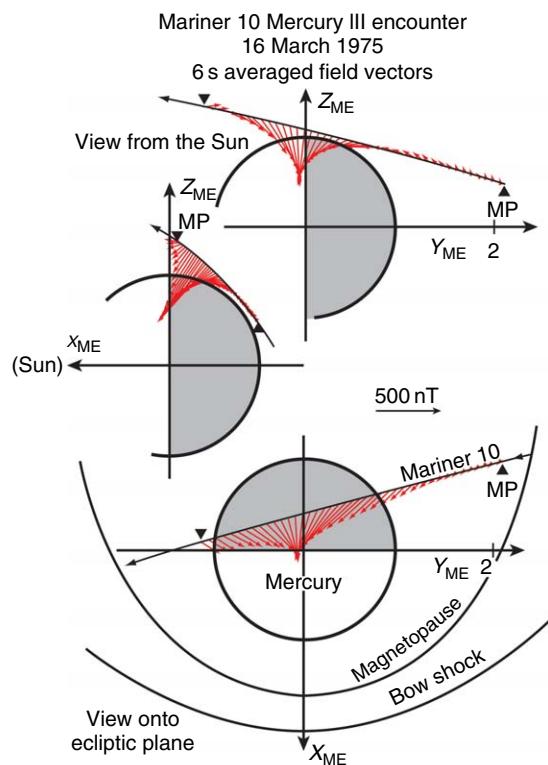


Figure 3 M3 encounter trajectory and magnetic field observations in Mercury ecliptic coordinates, viewed from the Sun (top) and viewed along Mercury's rotation axis (bottom). The vector magnetic field observed along the Mariner 10 trajectory is illustrated by vector projection onto the z-y (top), x-z (middle), and x-y (bottom) planes; vectors originate at the spacecraft position in ME coordinates at the time of observation. Observed magnetopause (MP) positions are shown along with scaled nominal bow shock and magnetopause boundaries.

orientation characteristic of a planetary magnetotail. As Mariner 10 crossed the equator plane, the field decreased in magnitude precipitously, as if crossing a current sheet, after which it was much smaller in magnitude and chaotic, lacking the smooth variation and consistent orientation that characterized the inbound pass. These observations indicated that Mariner 10 passed through a small magnetosphere, sized just large enough to contain the solid body of the planet under nominal solar wind conditions (Ness *et al.*, 1975). To form such a magnetosphere, Mercury must have an internal magnetic field, albeit one with a modest dipole magnitude.

Mariner 10's second flyby was upstream of Mercury in the solar wind, and too distant to penetrate the small Hermean magnetosphere. However, the third flyby was targeted to pass closely over the northern pole (Figure 3), in an effort to gather measurements optimized for characterization of the Hermean field. During the spacecraft's brief traversal

of the magnetosphere (14 min) the field increased from about 20 nT to a maximum of 402 nT observed just after closest approach at $1.14R_m$ radial distance ($1R_m = 2439$ km). The approximate $1/r^3$ dependence of the observed field on spacecraft radial distance establishes the internal origin of the field, and initial analysis (Ness *et al.*, 1976) suggested a dipole moment of approximately $342 \text{ nT} \cdot R_m^3$, aligned with the $-Z$ axis (normal to the Hermean orbital plane).

10.07.3.1.2 Models

Models of Mercury's magnetic field fall largely into two groups, distinguished by their treatment of external fields. The first group used a spherical harmonic expansion for both internal and external fields and

Table 2 Mercury magnetic field models

Model ^a	g_1^0 (nT-R _m ³)	Data	Reference	Comments
I1	227	I(1/2)	Ness et al. (1974a)	No external field modeled
I1E2	350	I(1/2)	Ness et al. (1975)	External field model improves internal fit
I1E1	342 ± 15	III	Ness et al. (1976)	III data better suited for modeling internal field
I1E1	330 ± 18	III	Ness (1979a)	III subsets modeled to estimate errors

^aII denotes internal expansion of degree/order 1; En denotes external field expansion of degree/order n.

were limited to the dipole ($n=1$) terms for internal fields. Initial analyses used the inbound Mercury I observations through to the tail current sheet crossing (1/2), while later analyses benefited from the Mercury III observations obtained crossing the pole. The latter are better suited to internal field estimation by virtue of improved signal-to-noise magnitude relative to external fields. These are listed in **Table 2**.

Another set of models were obtained using a spherical harmonic expansion for the internal field, augmented by explicit models of the external field. These included more detailed models of the fields due to magnetopause and tail currents, the latter being of great importance in interpretation of the Mercury I encounter. An explicit model of the external field allows one to extend analyses through regions of local currents (e.g., tail current sheet), accommodating nonpotential fields more readily. A number of these models (Whang, 1977; Jackson and Beard, 1977; Ng and Beard, 1979; Bergan and Engle, 1981) allowed for the possibility of internal field through degree 2 or 3, but only in the zonal harmonic ($m=0$) coefficient, for example, and g_2^0 and g_3^0 . Models allowing for axisymmetric quadrupole ($n=2$) or octupole ($n=3$) terms are characterized by a reduced dipole magnitude, relative to models allowing dipole terms only. This applies to all such models, regardless of author or method of analysis or even the details of the explicit models chosen to represent fields of external origin. This curious effect was shown to be related to model nonuniqueness, a consequence of the limited spatial extent of the data (Connerney and Ness, 1988). At the present time, that is, pending arrival of Messenger, there is no evidence for (or against) a nonzero axisymmetric quadrupole or octupole. Until then, a simple dipole model (e.g., Ness, 1979a) provides the best description of Mercury's magnetic field. The internal field is just sufficient to keep the dayside magnetopause above the planet's surface under nominal solar wind conditions (**Figure 4**).

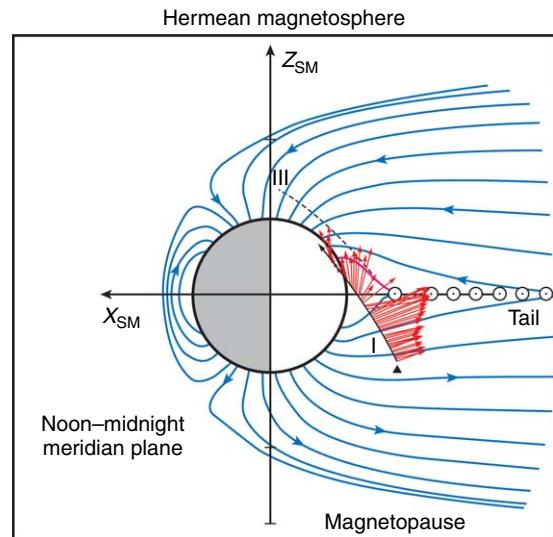


Figure 4 Magnetic field lines of a model Hermean magnetosphere (blue) in meridian plane projection (x-z plane) compared with the vector magnetic field observed along the M1 passage through the magnetotail.

10.07.3.2 Venus

Venus is similar to Earth in size ($R_v = 6051$ km), mass ($M_v = 0.815 M_e$), and density, and probably composition as well. Venus orbits the Sun once every 225 days at a mean orbital radius of 0.723 AU. Proximity to Earth has invited a relatively large number of missions, particularly in the early decades of space exploration (Colin, 1983; Russell and Vaisberg, 1983). However, unlike Earth, Venus has no measureable internal magnetic field; spacecraft have detected only weak magnetic fields associated with the solar wind interaction (Russell and Vaisberg, 1983). Venus also rotates very slowly, with a period of 243 days. The relatively slow rotation in itself does not explain the lack of a dynamo, as it is easily enough to make Coriolis forces dynamically important (Stevenson, 2003).

10.07.3.2.1 Observations

The first measurements of the magnetic field near Venus were obtained in December 1962, by the Mariner 2 spacecraft, on a flyby trajectory with a closest approach of 6.6 planet radii. Magnetic field measurements obtained by the triaxial fluxgate magnetometer (FGM) aboard Mariner 2 provided no evidence of an intrinsic planetary field (Smith *et al.*, 1963; Smith *et al.*, 1965) at this distance. Venus could not have an internal magnetic field with a moment exceeding 0.05 that of the Earth.

The 1960s and 1970s saw the arrival of the Venera 4 (1967) and Venera 6 (1969) entry probes and the Mariner 5 (1967) and Mariner 10 (1974) spacecraft on fly by trajectories with close approach distances of 1.7 and $1.96 R_v$. These were followed by the Venera 9 and 10 orbiters in 1975 and the pioneer Venus Orbiter (PVO) in 1978. The Veneras were placed in highly elliptical orbits with an orbital period of approximately 48 h, periapsis at 1545 and 1665 km altitude, respectively, and apoapses of approximately $18.5 R_v$. Throughout this period the existence of an intrinsic magnetic field was debated (see Russell and Vaisberg, 1983) amid ever-decreasing estimates, or upper limits on, the Venus dipole moment. The PVO was placed in a highly elliptical orbit with an orbital period of 24 h, periapsis at 140–170 km altitude, and apoapsis at approximately $12 R_v$. This orbit geometry afforded observation of, and a better understanding of, the interaction of the solar wind with a planetary atmosphere. These observations led to an upper limit on the Venus dipole moment of 4×10^{-5} that of Earth's (Russell *et al.*, 1980), which was reduced further (to $\sim 1 \times 10^{-5} M_e$) by a more exhaustive analysis (Phillips and Russell, 1987) that also demonstrated a lack of evidence for magnetic fields associated with crustal remanence.

10.07.3.2.2 Discussion

One concludes that Venus does not currently have an active dynamo; at present, little can be said of the possibility that Venus once sustained a dynamo. One would like to understand why this Earth-like planet does not have an active dynamo, and there are many possibilities (Stevenson, 2003). The favored explanation is that sufficient convection in the fluid outer core is lacking at present, owing perhaps to the inefficiency of heat transport (relative to Earth, where plate tectonics helps cool the core, see Chapter 10.09).

10.07.3.3 Mars

Mars is about half the size ($R_m = 3394$ km) of Earth and much less massive ($M_m = 0.108 M_e$). Mars resides at a mean orbital distance of 1.52 AU, orbiting the Sun once in 687 days. It has also been an attractive target for planetary exploration, playing host to more than 30 attempted missions since 1960 (Connerney *et al.*, 2004). At present, the Mars Exploration Rovers (MERs) continue to explore the surface (Crisp *et al.*, 2003), while MGS (Albee *et al.*, 1998), Mars Odyssey (Saunders *et al.*, 2004), Mars Express (Chicarro *et al.*, 2004), and the Mars Reconnaissance Orbiter (Graf *et al.*, 2005) all continue to acquire observations from orbit. It is the most impressive armada of active spacecraft ever assembled for planetary exploration.

10.07.3.3.1 Observations

After five unsuccessful attempts (four by the USSR; one by the USA), Mariner 4 became the first mission to return data from Mars on 15 July 1965. Mariner 4's fly by trajectory brought the spacecraft to within $4 R_m$ of the planet, from which vantage point it observed the near-Mars environment. The magnetic field was indistinguishable from the interplanetary environment; on bow shock, no magnetopause, no magnetosphere was observed. If Mars had an intrinsic magnetic field, it had to be very small in magnitude, less than $3 \times 10^{-4} M_e$ (M_e , the magnetic dipole moment of Earth, is $8 \times 10^{25} \text{ G cm}^3$), to remain unseen (Smith *et al.*, 1965). This upper limit is equivalent to an equatorial surface field of 100 nT, well less than Earth's (30 000 nT) but not so much less than the as yet undiscovered magnetic field of Mercury (330 nT).

In the following three decades, numerous spacecraft visited Mars, but no spacecraft, probe, or lander instrumented to measure magnetic fields would pass close enough to the surface to establish the presence of an intrinsic magnetic field. The US missions following Mariner 4, Mariner 9 (1971), and Viking 1 and 2 (1975), were not instrumented to measure magnetic fields. Additional magnetic field observations were obtained by the Soviet Mars 2, 3, and 5 spacecraft (Russell, 1979; Ness, 1979a) in the early 1970s and the Soviet Phobos 2 mission in 1988. These observations fueled a great deal of debate regarding the existence of a putative Mars dipole (with equatorial surface field in the range 20–65 nT) but no convincing evidence of an intrinsic field was found (Riedler *et al.*, 1989).

Thus it was the MGS spacecraft (Albee *et al.*, 1998), launched in 1996 as a (partial) replacement for the unsuccessful Mars Observer, which provided the first unambiguous detection of the intrinsic magnetic field of Mars. On 11 September 1997, the MGS spacecraft was inserted into a highly elliptical polar orbit from which it would transition to a circular polar mapping orbit (nominal 400 km altitude) via a series of aerobraking maneuvers (Albee *et al.*, 1998). Aerobraking uses atmospheric drag to slow the spacecraft, reducing apoapsis by repeated passage through the atmosphere; trim maneuvers are used to raise or lower the altitude at closest approach (periapsis) to obtain the desired drag. Each aerobraking pass brought the spacecraft over the surface of Mars, to altitude <100 km, about as close to the surface as any orbiting platform can achieve. During each aerobraking pass, the MGS magnetometer/electron reflectometer (MAG-ER) investigation (Acuña *et al.*, 1992) acquired measurements of the vector magnetic field along the orbit track at varying altitudes above the surface.

Periapsis pass 6 (day 264, 1997) provided a wealth of scientific insight in a few minutes time: lacking a significant field of global scale, Mars has no active dynamo but must have had one in the past when the crust acquired intense remanent magnetization (Acuña *et al.*, 1998). These observations are illustrated in Figure 5, showing the spacecraft trajectory and vector

magnetic field measured at 3 s intervals along the trajectory. From this image one can see that the intrinsic field is not global in scale, as it would necessarily be if it were generated deep within the planet, for example, by dynamo action. Instead, the field is localized to a small part of the trajectory, with a characteristic length scale comparable to the altitude of observation. This is the signature of a strong and localized magnetic source in the crust of Mars, just beneath the spacecraft orbit track. This particular source produces a field of nearly 400 nT at ~110 km altitude, requiring a very large volume of intensely magnetized material, with a moment of $\sim 1.6 \times 10^{16} \text{ A m}^2$ (Acuña *et al.*, 1998).

The MAG-ER investigation on MGS acquired magnetic field observations on over 1000 aerobraking passes during the transition to mapping orbit. On each such pass the crustal field was sampled at low altitude along a limited ground track centered on the latitude and longitude of periapsis and extending approximately $\pm 25^\circ$ in latitude north and south of closest approach. Periapsis passes were distributed more or less randomly in longitude, as the orbit period was shortened by aerobraking, and over the course of aerobraking, the latitude of periapsis slowly evolved from $\sim 30^\circ$ north latitude to $\sim 87^\circ$ south. Thus MGS was able to sample crustal fields over nearly the entire planet surface, which led to a map of the global distribution of magnetic sources in the Mars crust (Acuña *et al.*, 1999). This map, reproduced here (Figure 6), is compiled by color coding the (largely north-south) subspacecraft trajectory using the radial magnetic field component measured at that latitude and longitude. No correction for spacecraft altitude, which varies along the track from periapsis (~ 100 km) to 200 km, is attempted. The most significant magnetic sources are distributed nonrandomly, largely confined to the older, heavily cratered terrain south of the dichotomy boundary; vast regions of the smooth northern plains are relatively nonmagnetic, as are the Argyre and Hellas impact craters and prominent volcanic edifices (Acuña *et al.*, 1999). This led Acuna *et al.* to conclude that the Mars dynamo likely operated for a brief (few 100 My) period following accretion (~ 4.5 Ga) and had ceased to operate by the time of the late large impacts, about 3.9 Ga.

In March of 1999, MGS completed the transition to circular, polar mapping orbit with nominal altitude of 400 km, from which it would begin its mapping mission. The orbit is fixed in local time (2 a.m.–2 p.m.), and subsequent orbits cross the equator 28.6° westward of the previous orbit. Each ‘mapping cycle’ of 28 days duration provides uniform sampling of the globe with track-to-track separation of about 1° (59 km) at the

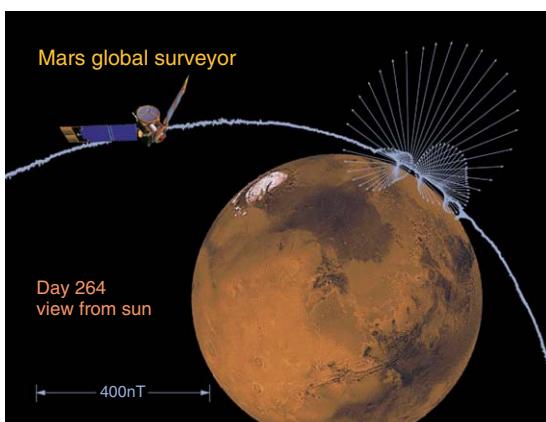


Figure 5 Mars Global Surveyor trajectory and magnetic field observations in an orbit plane projection for day 264, 1997. The vector magnetic field observed along the MGS trajectory is illustrated by vector projection onto the orbit plane; vectors originate at the spacecraft position at the time of observation. From Connerney JEP, Acuña MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Reviews* 111(1–2): 1–32.

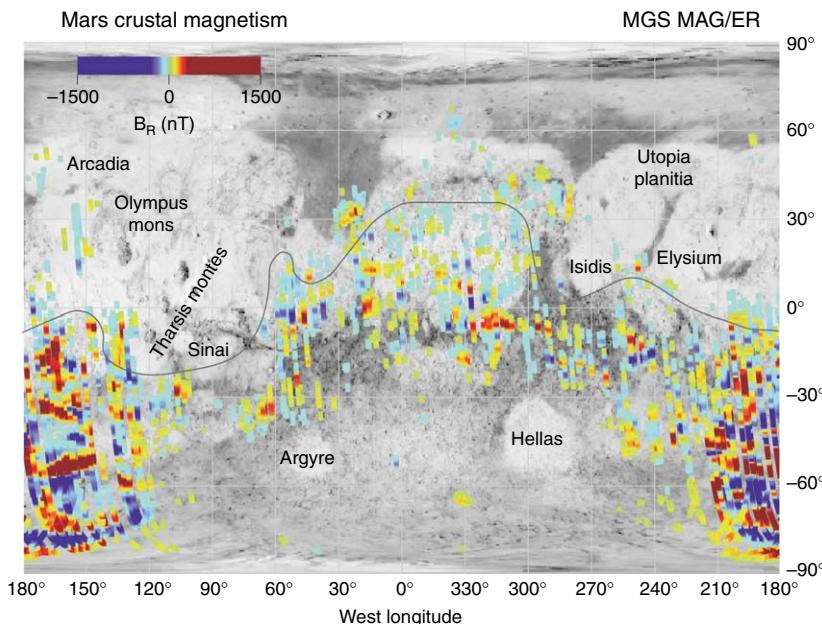


Figure 6 Global distribution of the most intense magnetic sources in the Mars crust. Magnitude of the radial magnetic field measured during the Mars Global Surveyor or aerobraking passes below 200 km altitude (to ~ 100 km minimum altitude). Most intense sources are found in the heavily cratered southern highlands, south of and near the dichotomy boundary (thin line). From Acuña MH, Connerney JEP, Ness NF, et al (1999) Global distribution of crustal magnetism discovered by the Mars Global Surveyor MAG/ER Experiment. *Science* 284: 790–793.

equator. Subsequent cycles increase the density of ground tracks, providing complete global coverage with track-to-track spacing of 3 km at the end of 1 Mars year (687 days). Thus MGS passes over the same position on Mars many times; the vector magnetic field at satellite altitude can be mapped with extraordinary signal-to-noise (Connerney *et al.*, 2001). This map was compiled with observations obtained over the nightside, where time-variable fields due to the solar wind interaction are minimized, to improve the signal fidelity of the crustal field. The crustal field at mapping altitude reaches a maximum of ~ 220 nT over the intensely magnetized southern highlands. This and other compilations of MGS observations are available online.

A much improved map of the field produced by crustal sources was obtained by using a more effective means to remove residual external fields (Connerney *et al.*, 2005). This map (Figure 7) is a 360×180 pixel ‘image’ of the filtered radial magnetic field, each pixel representing the median value in a $1^\circ \times 1^\circ$ latitude-longitude bin. This map represents the change in radial field, or ΔB_r , in nT per degree latitude, as the spacecraft moves north to south over the darkened hemisphere. The map is superposed upon a Mars Orbiter Laser Altimeter (MOLA) shaded topography map (Smith *et al.*, 1999), which appears in any pixel for

which ΔB_r falls below a minimum threshold of ± 0.3 nT per degree latitude. The threshold is chosen well above the noise background, so little noise appears in the figure to obscure the context image. The extraordinary signal fidelity of this map can be appreciated by recognizing the pixels adjacent in longitude are statistically independent, by virtue of the way MGS orbits accumulate. The map spans over two orders of magnitude of signal dynamic range.

The electron reflectometer (ER) instrument of the MAG/ER investigation (Acuña *et al.*, 1992; Acuña *et al.*, 1998) provides another means of sensing, remotely, the magnitude of the magnetic field between the spacecraft and atmosphere (Lillis *et al.*, 2004; Mitchell *et al.*, 2007). The technique is similar to that used in lunar orbit to infer surface magnetic field magnitudes (Anderson *et al.*, 1975; Lin *et al.*, 1998; Mitchell *et al.*, 2007) from the properties of the (reflected) electron distribution; however, at Mars, the presence of an absorbing atmosphere must be taken into account (Lillis *et al.*, 2004). Application of this technique at Mars is particularly effective in sensing weaker fields at an altitude of about 170 km above the surface. A global map of field magnitudes at this altitude (Mitchell *et al.*, 2007), with a spatial resolution of ~ 150 km and using a minimum B threshold of

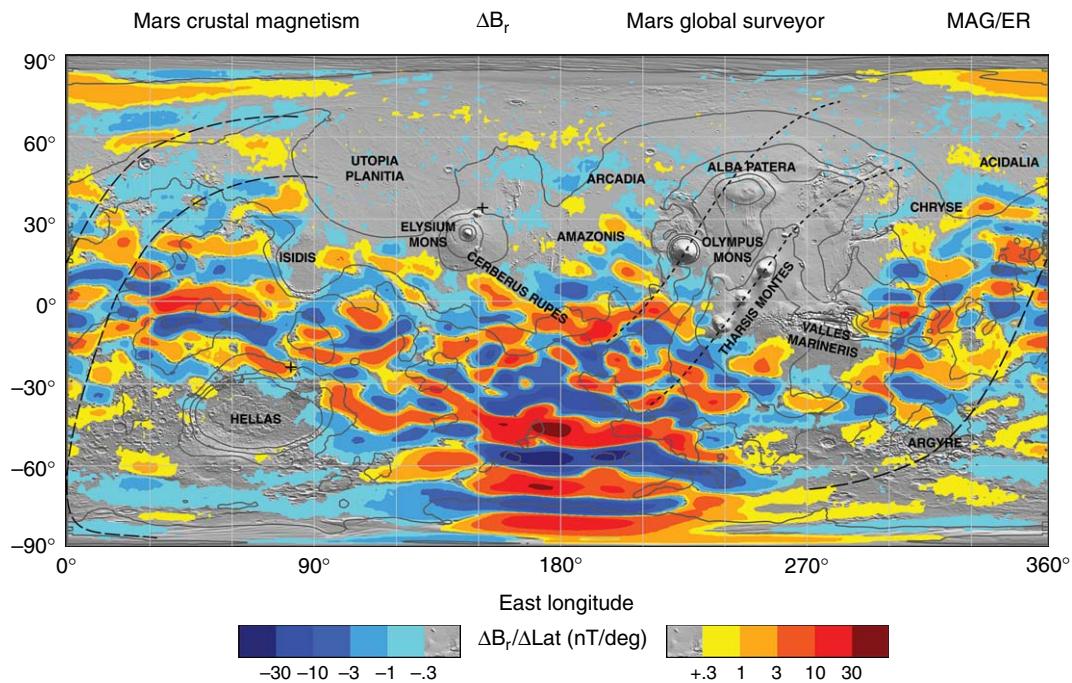


Figure 7 Map of the magnetic field of Mars observed by the Mars Global Surveyor satellite at a nominal 400 km (mapping) altitude. Each pixel is colored according to the median value of the filtered radial magnetic field component observed within the $1^\circ \times 1^\circ$ latitude/longitude range represented by the pixel. Colors are assigned in 12 steps spanning two-orders-of-magnitude variation. Where the field falls below the minimum value a shaded MOLA topography relief map (Smith et al., 1999) and contours of constant elevation ($-4, -2, 0, 2, 4$ km elevation) are shown. From Connerney JEP, Acuña MH, Ness NF, et al (2005) Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences* 102(42): 14970–14975.

10 nT at 170 km altitude, compares very well with the map (Figure 7) compiled from magnetic field measurements. The broad regions that appear in shaded topography in Figure 7, in which the magnetic field falls below threshold, are nearly identical to the weak-or zero-field regions (to an accuracy of a few nanoteslas) mapped by the ER.

10.07.3.3.2 Models

The representations routinely employed throughout the solar system – dipoles and spherical harmonic expansions – lose much of their appeal when applied to the crustal field on Mars. Such global models require an awkward number of model parameters or coefficients to represent the crustal field. For example, Cain's (Cain et al., 2003) spherical harmonic model extends to $N_{\max} = 90$, requiring $N_{\max} + 1)^2 - 1 = 8280$ model coefficients to calculate the field at one point above the surface. Likewise, Purucker's many-dipole model (Purucker et al., 2000) uses 11 500 dipoles distributed on a spherical surface ($1 R_m$) to represent the field by superposition. There are a variety of spherical harmonic (Cain et al., 2003; Arkani-Hamed, 2001a; Arkani-Hamed, 2002) and equivalent source models

(Purucker et al., 2000; Langlais et al., 2004) to choose from. These models can be obtained from the authors in electronic form. They offer a means of extrapolating the field from the measurement surface (400 km) and volume spanned by aerobraking passes to nearby points; however, neither can be expected to accurately predict the field near the source region.

Mars crustal magnetism is amenable to study using many of the methods and tools developed for interpretation of magnetic surveys on Earth. In 'Source modeling', one attempts to fit the observed vector field using one or more magnetized sources, for example, a collection of thin plates and prisms, often guided by constraints suggested by local geology and material properties. General techniques of 'continuation' of a potential field may also be applied to extrapolate the field away from a surface upon which the vector field is known. Downward continuation (toward the source) must be done with caution, as it is essentially a differencing operation that amplifies short-wavelength random noise (exponentially in z). Source modeling must also be done with caution, as it is not possible to uniquely determine crustal magnetizations using measurements external to the source.

Connerney *et al.*, (1999) used a source model to fit observations obtained during several aerobraking passes over the intensely magnetized southern highlands (Terra Cimmeria and Terra Sirenum). These quasi-parallel features appear lineated in the east–west (cross-track) direction; many can be traced over 1000 km in length. The model consisted of multiple uniformly magnetized parallel thin plates, aligned (and extending to infinity) in the east–west direction. The volume magnetization of each strip was determined using a linear inverse methodology to best fit the vector observations along one or more aerobraking passes (Figure 8). The observations could be fit well with a model characterized by strips of alternating polarity and volume magnetizations of as much as $\pm 20 \text{ A m}^{-1}$ (assuming a 30 km-thick plate). This figure is likely a lower bound, since only the product of thickness and volume magnetization is constrained by the observations (Connerney *et al.*, 1999). If the magnetization is borne in a thin layer, say 3 km, then volume magnetizations of $\pm 200 \text{ A m}^{-1}$ would be required. Connerney *et al.*, (1999) found the magnetic lineations on Mars sufficiently reminiscent of the pattern of magnetization associated with seafloor spreading on Earth that they proposed a similar mechanism – crustal spreading in the presence of a reversing dynamo – for the evolution of the Mars crust.

Spenke and Baker (2000) modeled crustal magnetization in the southern highlands using the same aerobraking observations and inverse methods used by Connerney *et al.* (1999). Their models imposed constraints on the direction of magnetization ('normal' and 'reversed') and allowed for a variation in the intensity of magnetization in the cross-track direction. They found similar magnetization intensities and found a good fit to the observations with a seafloor spreading model, allowing for variation of magnetization intensity (but not direction) along the stripes. The more uniformly distributed observations obtained at mapping altitude over the same region were downward continued using the Fourier transform method (Jurdy and Stefanick, 2004). Continuation to a 100 km-altitude surface allowed a favorable comparison with the aerobraking observations, and revealed abrupt lateral changes in field direction and magnitude suggestive of, but not requiring, reversed lineations.

There is great interest in identifying a Mars paleopole, and a number of authors have sought to determine ancient pole position(s) on Mars by fitting various magnetized sources to MGS magnetic field observations (Hood and Zakharian, 2001; Arkani-Hamed, 2001b; Spenke and Baker, 2000;

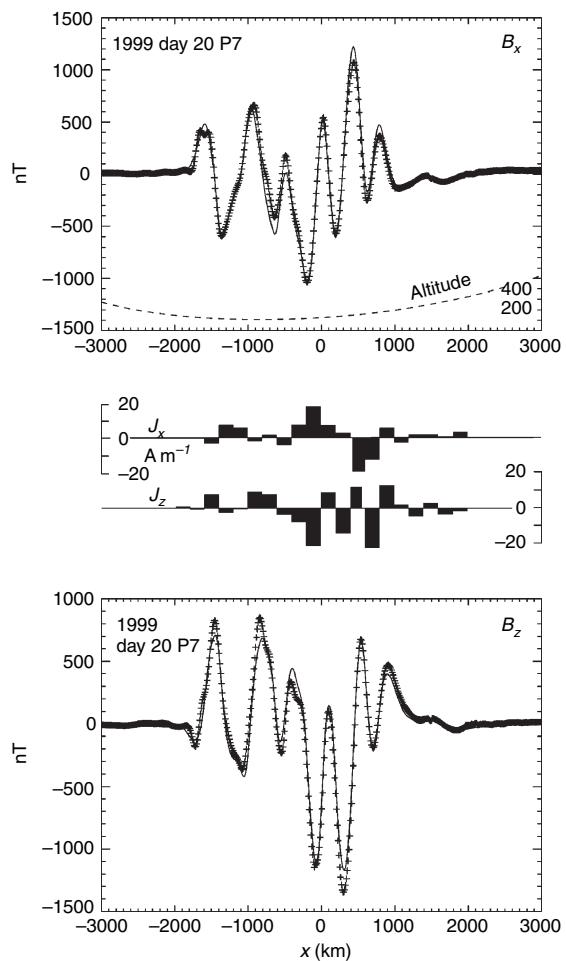


Figure 8 Vector magnetic field measured by MGS during aerobraking on calendar day 20, 1999. Periapsis occurred at 68.0°S , 181.2°W , and 106 km altitude. The x (north) and z (down) components of the vector field sampled every 3 s (crosses) are compared with a model fit (solid line). Altitude variation in kilometers is indicated with a dashed line. Observations plotted as a function of distance x north and south of an origin at 53°S and at the longitude of periapsis. Model consists of 20 uniformly magnetized slabs aligned with the y -axis (east–west) and infinite in extent along y . The x and z -components of the model crustal magnetization per unit volume (A m^{-1}) are indicated in the bar graphs between the two panels. From Connerney JEP, Acuña MH, Wasilewski PJ, *et al.* (1999) Magnetic lineations in the ancient crust of Mars. *Science* 287: 794–798.

Frawley and Taylor, 2004). A large number of putative paleopoles have been proposed but as yet no consensus has emerged. A general consideration of nonuniqueness in the interpretation of potential fields would suggest that without additional constraints the problem remains intractable (Connerney *et al.*, 2004). An alternative

approach, based on a statistical analysis of many sources (Sprenke, 2005), suggests a preference for paleopoles near 17° N, 230° E, and 17° S, 50° E, near those proposed earlier based on geomorphology and analysis of grazing impacts (Schultz and Lutz, 1988).

10.07.3.3.3 Discussion

Mars has no global field, therefore no dynamo at present, but must have had one in the past when the crust acquired intense remanent magnetization. (*see also* Chapter 10.09) It is likely that a molten iron core formed early, after, or during hot accretion 4.5–4.6 Ga, and for at least a few hundred million years, a substantial global field was generated by dynamo action in the core. The chronology proposed by Acuña *et al.* (1999) attributes the global distribution of magnetization to an early demise of the dynamo, prior to the last great impacts (~4 Ga). Schubert *et al.* (2000) argued in favor of a late onset of the dynamo, and propose that the southern highlands crust acquired magnetization from local heating and cooling events that postdate the era of large impacts and basin formation. Early onset and cessation of the dynamo is difficult to reconcile with the notion of a dynamo driven by solidification of an inner core (Schubert *et al.*, 1992), the preferred energy source for the Earth's dynamo. Alternatively, an early dynamo can be driven by thermal convection, with or without plate tectonics, for the first 0.5–1 Gy (Breuer and Spohn, 2003; Schubert and Spohn, 1990; Stevenson *et al.*, 1983; Connerney *et al.*, 2004), persisting as long as the core heat flow remains above a critical threshold for thermal convection (Nimmo and Stevenson, 2000).

After more than 2 full years of mapping operations, MGS has produced an unprecedented global map of magnetic fields produced by remanent magnetism in the crust (Connerney *et al.*, 2005). This map (Figure 7) reveals contrasts in magnetization that appear in association with known faults; variations in magnetization that are clearly associated with volcanic provinces; and magnetic field patterns that appear shifted along small circles in the manner of transform faults at spreading centers (Connerney *et al.*, 2005). Connerney *et al.* proposed that the entire crust acquired a magnetic imprint via spreading in the presence of a reversing dynamo and that erasure of this imprint occurred where the crust was buried (thermal demagnetization) by flood basalts to depths of a few kilometers. In the remaining magnetic imprint one can identify the tell-tale signature of transform faults, in Meridiani, and the signature of craters that formed both before and after the demise of the dynamo. Of course, transform faults

are unique to plate tectonics, so if these features are indeed transform faults then the Mars crust formed via seafloor spreading as on the Earth (Connerney *et al.*, 1999; Sleep, 1990). The interpretation of this map is not without controversy, however, and one may expect much more from these observations as the analyses mature.

Magnetic mapping is an even more powerful tool on Mars than on Earth, where it has been indispensable in understanding the evolution of the Earth, especially in the development of the unifying theory of plate tectonics. On Earth, magnetic surveys are complicated by the presence of the main field, which makes measurement of crustal anomalies challenging (limited signal to noise) and more difficult to interpret (induced or remanent field?). Mars is free of such complications; the global map of its crustal field is unique, affording an opportunity to apply the methods of exploration geophysics on a global scale. It is clear that the crust of Mars retains a magnetic imprint of its formation and subsequent evolution. That history may be read with increasing resolution as magnetic surveys are conducted at lower altitudes. One may anticipate great insights from such surveys as well as Mars paleomagnetism when that field develops.

10.07.4 Gas Giants

10.07.4.1 Jupiter

Jupiter, the largest ($R_j = 71\,372$ km) and most massive ($M_j = 318\,M_e$) planet in the solar system, resides at a mean orbital distance of 5.2 AU from the Sun. Jupiter, capturing much of the gas component of the solar nebula, is largely a solar mix of H and He, possibly with a rock and ice core of some tens of Earth masses. The Jovian system (*see also* Chapter 10.13 for a discussion of the giant planets) has been visited often by spacecraft, both as a destination and as a waypoint for vessels destined to travel onward. A passing spacecraft can use Jupiter's considerable mass and orbital velocity to advantage, acquiring a boost to greater radial distances (e.g., Cassini, New Horizons to Pluto) or to achieve higher solar latitudes (e.g., Ulysses), than would otherwise be possible. Thus, Jupiter's orbit serves as a crossroads for travelers to other worlds as well.

Jupiter was known to have a magnetic field long before the first *in situ* observations were obtained. Burke and Franklin (1955) had detected nonthermal decameter-wavelength emissions (22 MHz) from

Jupiter, leading to speculation that Jupiter possessed a magnetic field of internal origin. Observations at decimetric (~ 1 GHz) wavelengths subsequently revealed synchrotron radiation, emitted by high-energy electrons trapped in a Jovian Van Allen belt. This emission, as observed, is modulated by the rotation period of the planet (9.925 h). Since synchrotron radiation is narrowly beamed in a direction perpendicular to the magnetic field guiding the electron motion, the modulation can be used to infer the geometrical properties of the field. Thus, the tilt (9.5°) and phase ($200^\circ \lambda_{III}$) of the Jovian dipole were deduced (Morris and Berge, 1962), and estimates of the dipole offset (from the origin) were made (Warwick, 1963), well before the first spacecraft arrived at Jupiter. The magnitude of the field, however, remained in doubt.

Continued observations of Jovian radio emission at decameter wavelength revealed yet another periodicity in the occurrence of radio storms, one that matched the orbital period (42.46 h) of the innermost Galilean satellite, Io (Bigg, 1964). The intensity and occurrence rate of radio storms varies dramatically with Io's orbital phase, as well as Jovian central meridian longitude (CML). Activity peaks when Io's orbital phase is near 90° and 240° (measured in the

direction of Io's orbital motion from geocentric superior conjunction). Io's influence on Jovian radio emissions was attributed to an electrodynamic interaction (Goldreich and Lynden-Bell, 1969) that drives currents along magnetic field lines to and from Jupiter's magnetosphere.

10.07.4.2 Observations

Pioneer 10, the first spacecraft to enter the Jovian magnetosphere, passed within $2.8 R_J$ of the planet at closest approach in December 1973. Pioneer 10's flyby trajectory carried it through the Jovian system at relatively low magnetic latitudes (Figure 9). This region, near the equator, is home to a large-scale system of azimuthal currents (ring currents) extending from about $5 R_J$ to the outer reaches of the magnetosphere ($\sim 100 R_J$). This current system imparts a disk-like geometry to the magnetic field, with field lines at great distances stretched nearly parallel to the equator (leading to the descriptive 'Jovian magnetodisc'). Precisely 1 year later, Pioneer 11 approached to within $1.6 R_J$ on a retrograde λ_{III} trajectory. The Pioneer 11 trajectory provided measurements over a wide range of planet latitudes (Figure 9) and longitudes.

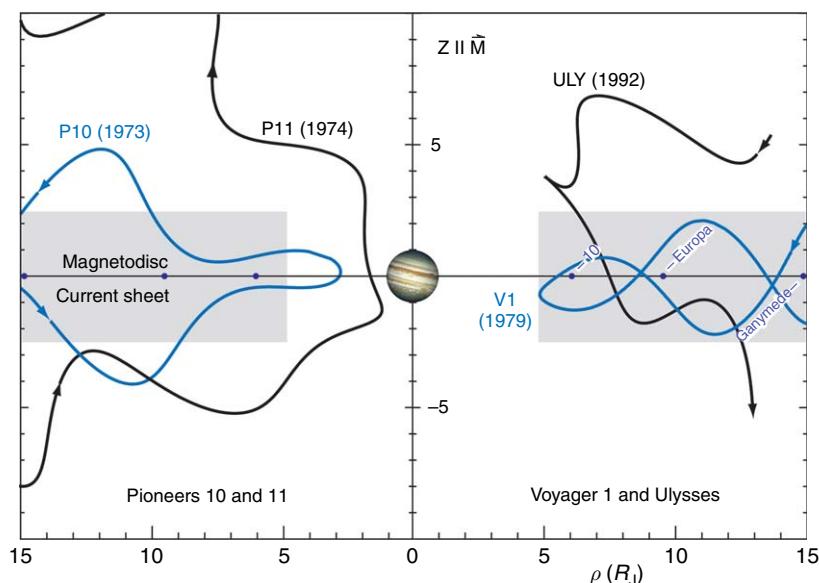


Figure 9 Spacecraft flybys (Pioneer 10 and 11, Voyager 1, and Ulysses) of Jupiter in a magnetic equatorial (cylindrical) coordinate system. The z -axis is aligned with Jupiter's magnetic dipole axis. The spacecraft distance from the magnetic equator (z) as a function of distance from the z -axis (ρ) is given in units of Jupiter radius. The Galilean satellites Io, Europa, and Ganymede trace out arcs (not shown) in this coordinate system crossing the equator at the radial distances indicated. The shaded region approximates the washer-shaped system of azimuthal ring currents that gives rise to the magnetodisc configuration of the Jovian magnetosphere.

The Pioneer 10 and 11 spacecraft were instrumented to measure magnetic fields with a vector helium magnetometer (VHM) provided by the Jet Propulsion Laboratory (Smith *et al.*, 1975). This instrument could measure a maximum of ± 1.4 G in the highest of its eight dynamic ranges. Pioneer 11 also carried a small high-field triaxial FGM provided by the Goddard Space Flight Center (Acuña and Ness, 1976). The FGM was added as a precaution, to extend measurement capability to a field of 10 G along each axis (Earth's field magnitude ~ 0.6 G at the pole). The magnitude of Jupiter's field was unknown at the time of launch, and the concern was that without a high-field instrument Pioneer 11 might not have the dynamic range needed to make observations throughout the encounter. Fortunately, the maximum field experienced by Pioneer 11 was ~ 1.12 G, observed just as the spacecraft went into occultation on its way to closest approach at $1.6 R_j$. There are small but significant differences between the two data sets that have not been fully resolved, so most analyses have tended to use data from one or the other but not both.

The two Voyager encounters followed in March and July of 1979. Like Pioneer 10, the Voyagers remained near the Jovigraphic equator, confined to relatively low magnetic latitudes. Voyager 1 was targeted for a close flyby of Io and the Io flux tube (IFT); its closest approach to Jupiter was $4.9 R_j$, just inside the orbit of Io ($5.95 R_j$). The maximum field measured by Voyager 1 was just ~ 3330 nT. The Voyager 2 flyby was relatively distant, with a closest approach distance of $10R_j$. In this region of the magnetosphere, the measured magnetic field has a significant contribution due to the magnetodisc currents. These will tend to reduce the θ component of the planetary field near the equator and introduce a component of the field in the radial direction which changes sign in crossing the magnetic equator. Local currents cannot be modeled with potential fields and if present, may lead to large errors in estimates of the planetary field if not accounted for (Connerney, 1981).

Jupiter was not visited again until the Ulysses flyby on 8 February 1992, more than two decades after the Voyager flybys. Ulysses was designed to study the heliosphere at high solar latitudes; the Jupiter flyby was used to kick the spacecraft into a high-inclination orbit about the Sun. The spacecraft approached Jupiter at relatively high northern latitude, measuring a maximum field of 2372 nT just after closest approach at $6.3 R_j$. This encounter was

much like that of Voyager 1, in that for much of the time nearest closest approach the spacecraft was immersed in the local currents of the Jovian magnetodisc.

The recently completed Galileo mission to the Jovian system was designed to deliver an atmospheric probe, explore the distant magnetosphere, and provide opportunities for many satellite encounters (Johnson *et al.*, 1992). It did not provide appreciable additional information on the planetary field. Likewise, the Cassini spacecraft encounter was a distant flyby, engineered to speed the spacecraft onward to its destination (Saturn) in the outer solar system.

10.07.4.3 Models

The large number of Jovian magnetic field models available is a result of several factors, interest in the Jovian system undoubtably foremost among them. There is a wealth of observations of Jovian phenomena, many of which (e.g., radio emission, aurorae) depend critically on details of the magnetic field near the planet where few direct observations exist. The need is great. In addition, techniques for modeling the field have evolved, and improved, in part in response to the challenges of separating the planetary field from that produced by external currents. Techniques have also evolved to take advantage of new observations (e.g., emission from the foot of the IFT) that serve as valuable constraints on the field near the surface. Here we describe a subset of the available models, intended to illustrate the development of new approaches.

The first models of Jupiter's planetary field were the OTD approximations (D2 and D4) obtained from the Pioneer 10 and 11 VHM observations (Smith *et al.*, 1976). Analyses of the Pioneer 11 FGM observations suggested that a more capable parametrization was needed, leading to development of the O₄ model (Acuña and Ness, 1976), the first of many spherical harmonic models of Jupiter's internal field. This model and the SHA model of Smith *et al.* (1976) allowed spherical harmonics of degree and order 3 (octupole) for the internal field and approximated external fields with external harmonics of degree 1 or 2. **Table 3** gathers together the spherical harmonic coefficients of several models, fit to various subsets of data and using a variety of methods. Estimated errors are given for some of these models, but not reproduced here, for brevity. The original publication should be consulted.

Table 3 Jovian magnetic field models: spherical harmonics

Coefficient Number	Coefficient	VIT 4:	VIP 4:	1992.1: Ulysses 17ev		O ₆	1979.2: Voyager 1 17ev		1979.2: Pioneer 11	
				O ₄	SHA		O ₄	SHA	O ₄	SHA
1	g_1^0	428077.	420543.	410879.	424202.	420825.	421800.	409200.		
2	g_1^1	-75306.	-65920.	-67885.	-65929.	-65980.	-66400.	-70500.		
3	h_1^1	24616.	24992.	22881.	24116	26122.	26400.	23100.		
4	g_2^0	-4283.	-5118.	7086.	-2181.	-3411.	-20300.	-3300.		
5	g_2^1	-59426.	-61904.	-64371.	-71106.	-75856.	-73500.	-69900.		
6	g_2^2	44386.	49690.	46437.	48714.	48321.	51300.	53700.		
7	h_2^1	-50154.	-36052.	-30924.	-40304.	-29424.	-46900.	-53100.		
8	h_2^2	38452.	5250.	13288.	7179.	10704.	8800.	7400.		
9	g_3^0	8906.	-1576.	-5104.	7565.	2153.	-23300.	11300.		
10	g_3^1	-21447.	-52036.	-15682.	-15493.	-3295.	-7600.	-58500.		
11	g_3^2	21130.	24386.	25148.	19775.	26315.	16800.	28300.		
12	g_3^3	-1190.	-17597.	-4253.	-17958.	-6905.	-23100.	6700.		
13	h_3^1	-17187.	-8804.	-15040.	-38824.	8883.	-58000.	-42300.		
14	h_3^2	40667.	40829.	45743.	34243.	69538.	48700.	12000.		
15	h_3^3	-35263.	-31586.	-21705.	-22439.	-24718.	-29400.	-17100.		
16	g_4^0	-22925.	-16758.							
17	g_4^1	18940.	22210.							
18	g_4^2	-3851.	-6074.							
19	g_4^3	9926.	-20243.							
20	g_4^4	1271.	6643.							
21	h_4^1	16088.	7557.							
22	h_4^2	11807.	40411.							
23	h_4^3	6195.	-16597.							
24	h_4^4	12641.	3866.							
<i>Magnetodisc</i>										
	R_0			7.1			5.(UR)			
	R_1			128. (UR)			56.			
	D			3.3			3.1			
	$\mu_0 l_0 / 2$			137.			185.			
	Θ_0			8.2			6.5			
	Φ			200.			206.			

Schmidt normalized spherical harmonic coefficient in gauss, referenced to Jupiter system III (1965) coordinates, and 1 $R_j = 71\,398$ km for Ulysses; 1 $R_j = 71\,323$ km for Voyager 1. Voyager 1 17ev model from Connerney *et al.* (1982a).

The notation 'UR' refers to unresolved parameters. Pioneer 11 O₄ model coefficients as tabulated for system III (1965) by Acuña *et al.* (1983b) (originally (1957 system III) from Acuña and Ness (1976)).

Pioneer 11 SHA model originally (1957 System III) from Smith *et al.* (1976).

The Pioneer 11 VHM and FGM observations have proved to be most useful for studies of the planetary field, in large part due to the relatively good spatial distribution of observations afforded by that spacecraft's trajectory. The problem of inverting flyby observations to obtain estimates of the spherical harmonic coefficients has been posed as an eigenvalue problem and solved using the SVD method (Connerney, 1981). This technique allows partial solutions to underconstrained linear inverse problems and can be used to show how even small measurement errors – such as noise or unmodeled

external fields – can lead to large errors among the internal field coefficients. Pioneer 11 also remained largely outside of the region of current flow (equatorial magnetodisc) near closest approach so external fields are less of a problem.

The Voyager 1 observations yielded a considerably reduced dipole moment, upon preliminary analysis, which was attributed to the presence of magnetodisc currents that were not adequately modeled (Ness *et al.*, 1979a). Subsequent analyses of the Voyager 1 observations used an explicit model of the magnetodisc (Connerney *et al.*, 1981) in combination

with a spherical harmonic model for the planetary field. An inversion scheme that allowed for variation of the parameters of the magnetodisc along with those of the internal field resulted in a partial octupole model (V1 17ev) of the planetary magnetic field (Connerney *et al.*, 1982a). The Voyager 1 model of the field (epoch 1979) could be compared with the Pioneer 11 model (GSFC O₄) of the field (epoch 1973) to limit Jovimagnetic secular variation (g_1^0) to no more than 0.2% yr⁻¹ (Connerney and Acuña, 1982). By comparison, the present-day decrease in the terrestrial g_1^0 term is about 0.075% yr⁻¹.

Preliminary analysis of the Ulysses encounter observations (Balogh *et al.*, 1992) suggested good agreement with the magnetic field calculated using the O₆ planetary field augmented with a magnetodisk current sheet model (Connerney *et al.*, 1981) if one allowed for variations of the external field during the encounter. The Ulysses close approach observations were obtained in the immediate vicinity of the inner edge of the current sheet, so some attention to external fields is required. Dougherty *et al.* (1996) found the data consistent with Connerney's O₆ model, allowing for a reduction of the g_1^0 coefficient (main dipole term), from 4.242 to 4.059 G. Connerney *et al.* (1996) derived an octupole model of the field (Ulysses 17ev), also with a reduced dipole g_1^0 coefficient (4.109 G), and demonstrated that the internal zonal harmonic coefficients covary with the parameters of the current disk, complicating attempts to separate internal and external fields. The magnetodisk current system was considerably weaker during the Ulysses epoch, relative to that observed during the Pioneer and Voyager epochs. Connerney *et al.* (1996) quote a figure of 20% less total integrated magnetodisk current at the time of Ulysses encounter, Dougherty *et al.* (1996) quoting a 28% reduction. Differences among the Ulysses era (Ulysses 17ev) model and those of Voyager (Voyager 1 17ev) and Pioneer (O₄) appear within, or comparable, to estimated parameter errors (Connerney *et al.*, 1996); so there is little evidence of secular variation of the main field as yet.

The O₆ model of Jupiter's magnetic field used observations from both Pioneer 11 (1973) and Voyager 1 (1979), assuming that the magnetic field has not changed appreciably between encounters, and adds a constraint on Jupiter's harmonic spectrum (Connerney, 1992). This model was designed to benefit from the better spatial distribution of observations afforded by combining trajectories. It was also designed to allow for the possibility that unmodeled

fields of higher degree and order than octupole might be present in the observations (particularly those of Pioneer 11, passing to within 1.6 R_j). This model represents a partial solution to an underdetermined inverse problem (spherical harmonic expansion to degree and order 6) combined with an explicit model of the magnetodisc. Another distinguishing feature of this model was the application of weights $(r_c/a)^{n+2}$ to parameters of degree n , where r_c , the core radius, was taken to be 0.7 R_j . This weighting encourages a model solution with terms of degree n contributing equally to the mean-squared field amplitude on the (presumed) core boundary, as is observed for Earth, through degree 13 or so (Langel, 1987). The Schmidt coefficients of this model ('O₆' for octupole part of a degree 6 expansion) are listed in Table 3. The O₆ model is characterized by smaller quadrupole and octupole moments, relative to the comparable O₄ model of Acuña and Ness (1976), and in that regard is more Earth-like in harmonic content (Figure 10).

The discovery of infrared emission at the foot of the IFT in Jupiter's ionosphere (Connerney *et al.*, 1993) provided an entirely new and extremely valuable constraint on magnetic field models. This emission occurs in Jupiter's polar ionosphere where field lines threading the satellite Io (or its conducting ionosphere) at an orbital distance of 5.95 R_j intersect the Jovian ionosphere. The feature is observed to move across the disk of Jupiter in concert with Io's orbital motion, tracing a path around each magnetic pole (Figure 11). This path serves as a fiducial marker, providing an unambiguous reference on Jupiter's surface through which magnetic field lines with an equatorial crossing distance of 5.95 R_j must pass. Thus, observations of the location (latitude, longitude) of the footprint offer a unique constraint on magnetic field models, precisely where it is most needed, on the surface of the planet, and otherwise unavailable. This emission is observable from Earth with ground telescopes (e.g., Infrared Telescope Facility on Mauna Kea) in the infrared region of the spectrum (Connerney *et al.*, 1993; Connerney and Sato, 2000), and with the Hubble Space Telescope in the ultraviolet (Clarke *et al.*, 1995, 1996, 2005; Prange *et al.*, 1996). Emission has now been detected at the foot of the Europa and Ganymede flux tubes as well (Clarke *et al.*, 2002). These observations are less useful constraints on the internal field because these field lines pass through the equator at greater radial distance (9.4 and 15 R_j , respectively) where relatively large and variable external fields are encountered.

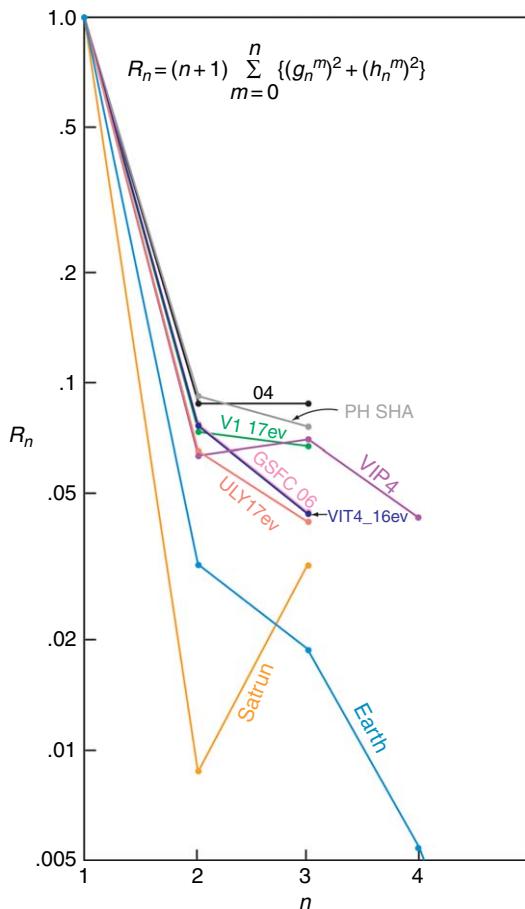


Figure 10 Relative harmonic content (Lowe's spectrum) of spherical harmonic models of the Jovian field, compared with that of Earth and Saturn.

Over 100 determinations of the location of the IFT footprint were combined with *in situ* magnetic field observations (Pioneer 11 VHM and Voyager 1 FGM) to obtain a partial solution to a fourth degree and order expansion of the internal field (Connerney *et al.*, 1998). This model is referred to as the 'VIP4' model (**Table 3**), reflecting use of Voyager 1, IFT, and Pioneer observations; and an internal spherical harmonic expansion to degree and order 4. This model was designed to fit the position of the IFT footprint very well (within $\sim 1^\circ$ latitude), so it has proven particularly useful in analyses of satellite interactions and aurorae. However, no one model fits the IFT footprints and 'all' of the *in situ* magnetic field observations to within their expected errors (Connerney *et al.*, 1998), although the IFT observations can be so fit with any 'one' set of flyby observations.

Surprisingly, the IFT footprint observations are nearly sufficient 'by themselves' to yield a useful field

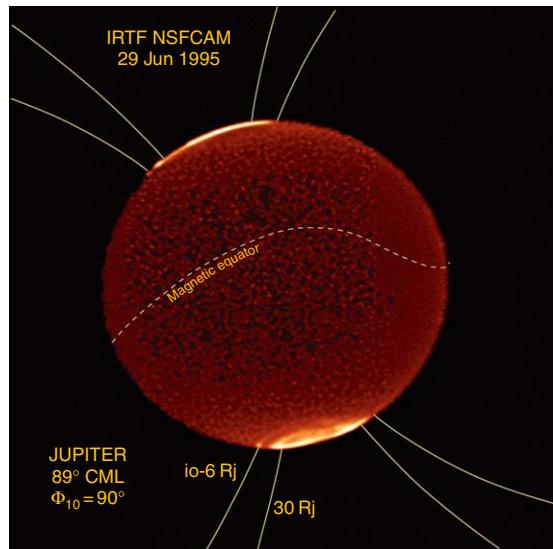


Figure 11 Image of Jupiter at $3.4\text{ }\mu\text{m}$ obtained at NASA's IRTF at Mauna Kea, Hawaii, using the NSFCAM facility imager. At this wavelength, bright H_3^+ auroral emissions originating above the homopause appear again as a planetary disc darkened by methane absorption. Magnetic field lines are drawn in meridian plane projection to illustrate where field lines crossing the equator at $30R_{\text{J}}$ and $6R_{\text{J}}$ (Io's orbit) intersect Jupiter's surface. A bright emission feature at the foot of the Io Flux Tube (IFT) can be identified near the dawn limb, as well as faint emission extending along the Io L-shell downstream of the instantaneous IFT. From Connerney JEP and Satoh T (2000) The H_3^+ ion: A remote diagnostic of the Jovian magnetosphere. *Philosophical Transactions of the Royal Society of London A* 358: 2471–2483.

model. All that is needed is a means of constraining the magnitude of the field. The final model in **Table 3** lists a model derived using over 500 IFT footprint locations (Connerney *et al.*, 1998). This model constrains the field magnitude using a minimum of *in situ* magnetic field data: the θ component, only, of the Voyager 1 closest approach data obtained within $7R_{\text{J}}$. The Voyager 1 data was chosen as the most accurate ($\sim 0.1\%$) available and the θ component, ranging from 1036 nT to a maximum of 3280 nT, as the least influenced by external fields (larger in magnitude than the radial component of the external field but easily modeled). This model ('VIT4', reflecting use of IFT observations, Voyager θ component, degree and order 4) is essentially a product of remote observation, using just enough *in situ* data to constrain the magnitude of the field. In principle, any constraint on the field magnitude – such as the field magnitude on the surface, deduced from the maximum frequency of radio emission at the electron cyclotron frequency – would serve as well.

10.07.5 Discussion

The magnitude of the surface field for the O_6 model ranges from just over 3 G at low latitudes to just over 14 G at high (northern) latitudes. This is similar to the range of field magnitudes predicted by the O_4 model of Acuña and Ness (1976). The variation of magnetic field magnitude on the surface of Jupiter is illustrated in **Figure 12**, which depicts contours of constant field

magnitude on the dynamically flattened surface of Jupiter, computed using the VIT4 model. This model has a less prominent north–south asymmetry in the magnetic field magnitudes, compared to earlier models; the maximum field in the Northern Hemisphere (13.9 G) is comparable to that of earlier models (see Acuña *et al.*, 1983b), but the maximum field in the Southern Hemisphere (13.2 G) is considerably larger. The estimated uncertainty in the surface field

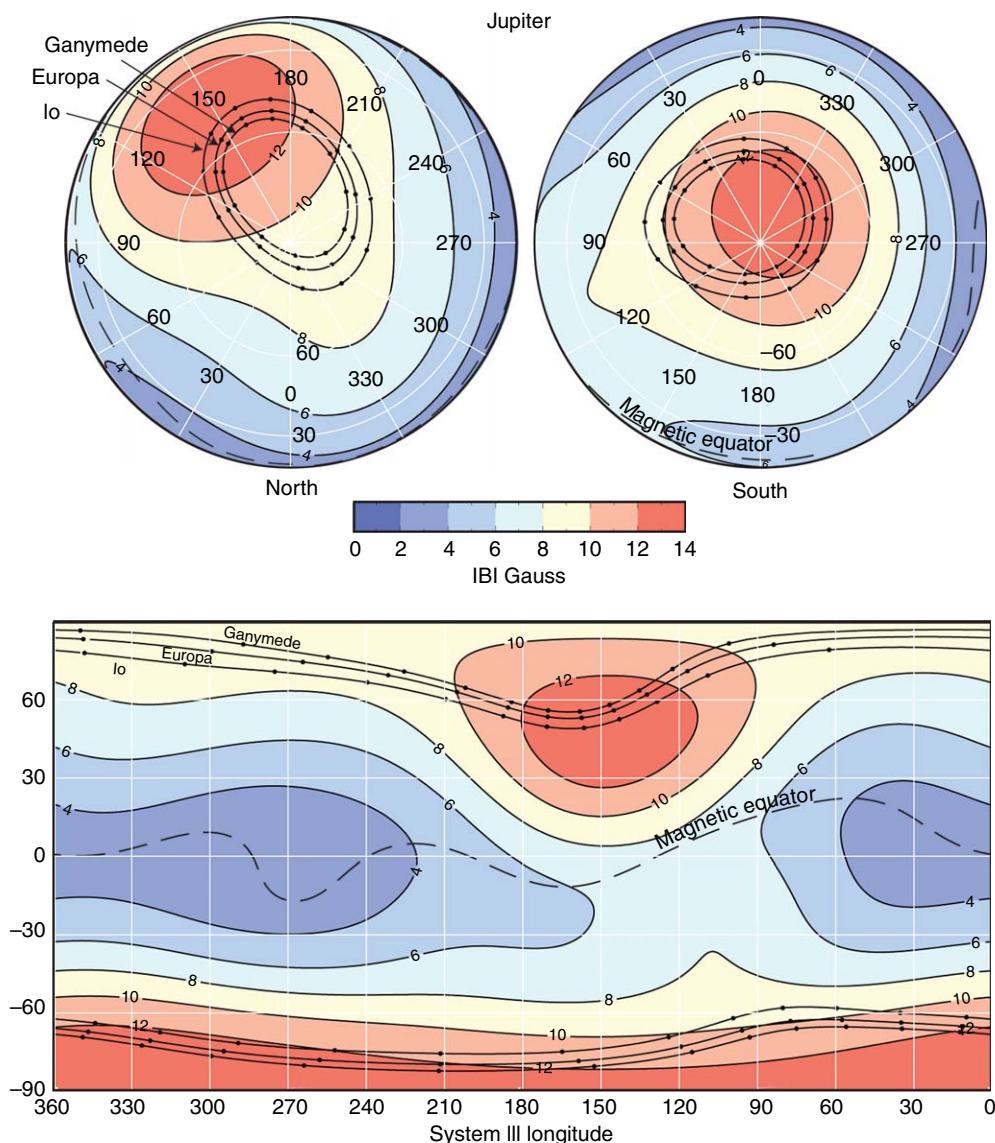


Figure 12 Contours of constant magnetic field magnitude (Gauss) on the dynamically flattened ($I/15.4$) surface of Jupiter, computed using the VIT4 model field (see text). The top panel shows the field magnitude in orthographic projections viewed from the north (left) and south (right); below the color bar is a rectangular latitude–longitude (System III west longitude) projection. A trace (dashed) indicates the position of the magnetic equator. Solid lines join points on the surface that trace along field lines to the orbits of the Galilean satellites Io, Europa, and Ganymede, with filled circles for increments of 30° in the satellite’s System III longitude.

magnitude is about ± 1 G (Connerney, 1981, 1992), but could be appreciably greater if higher-degree and -order harmonics are larger than expected. **Figure 12** also illustrates the foot of the IFT ('Io foot'), which is the path traced out on the planet's surface by the field lines which pass through the satellite Io as the planet rotates. The foot of the IFT passes through the region of highest field strength in the Northern Hemisphere at a longitude of about $150^\circ \lambda_{\text{III}}$. The maximum surface magnetic field magnitude present along the Io foot is consistent with the maximum frequency extent (39.5 MHz) of Jovian decameter radio emission (DAM), assuming that the emission occurs at the local electron gyrofrequency and at the foot of the IFT ($f_c(\text{MHz}) = 2.8 B (\text{G})$).

10.07.5.1 Saturn

Saturn, the second largest ($R_s = 60\,000$ km) and most massive ($M_s = 95 M_e$) planet in the solar system, resides at a mean orbital distance of 9.5 AU, nearly twice as far from the Sun as Jupiter. Like Jupiter, Saturn exhibits the characteristics of a condensed object of solar composition. Saturn was assumed to be much like its larger companion before Pioneer 11 arrived. However, unlike Jupiter, Saturn as a radio source was quiet. Saturn is too weak a radio source and too distant to be easily detected from Earth. In addition, Saturn emits at lower frequencies (a few

kilohertz to 1 MHz) than Jupiter, well below the critical frequency required to penetrate the Earth's ionosphere. Thus, Saturnian radio emissions cannot be monitored using terrestrial radio receivers.

10.07.5.1.1 Observations

The discovery of Saturn's magnetic field was thus left to Pioneer 11, arriving in September 1979, following its Jupiter swingby. The particle and fields investigations on broad Pioneer 11 obtained observations along a spacecraft trajectory that remained within approximately 6° of the equator (**Figure 13**). The most remarkable feature of Saturn's magnetic field proved to be the close alignment of the magnetic dipole and rotation axes (Smith *et al.*, 1980a; Acuña and Ness, 1980; Acuña *et al.*, 1980), also noted by the charged particle investigations (Simpson *et al.*, 1980; Van Allen *et al.*, 1980). Pioneer 11 approached to within $1.35 R_s$, measuring a maximum field of ~ 8200 nT at -6° latitude (Smith *et al.*, 1980a).

The Voyager 1 encounter in November 1980 and the Voyager 2 encounter in August 1981 provided the first observations of Saturn's magnetic field at high northern and southern latitudes (Ness *et al.*, 1981, 1982). Voyager 1 sampled relatively high latitudes (-40°) at close radial distance ($3.07 R_s$) but remained in the Southern Hemisphere while inside of $6 R_s$ radial distance. Voyager 1 measured a maximum field of 1093 nT at -40° latitude and 184° SLS longitude,

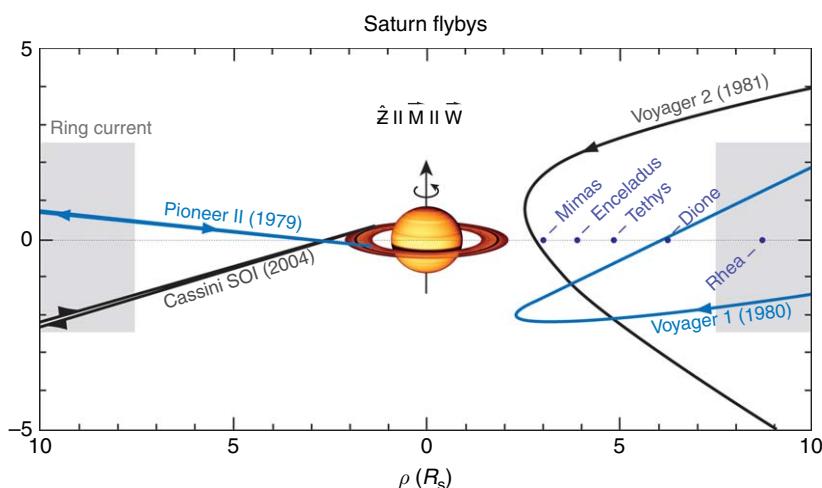


Figure 13 Spacecraft flybys (Pioneer 11, Voyager 1 and 2, and Cassini) of Saturn in a cylindrical coordinate system in which the z-axis is aligned with Saturn's magnetic dipole axis and rotation axis. The spacecraft distance from the equator plane (ρ) is given in units of Saturn radius. The positions of satellites Mimas, Enceladus, Tethys, Dione, and Rhea are indicated. The shaded region approximates the innermost portion of a washer-shaped region of azimuthal ring currents.

just prior to closest approach (Ness *et al.*, 1981). Voyager 2's closest approach of $2.69R_s$ occurred at 323° longitude, diametrically opposite those of Pioneer 11 and Voyager 1. Voyager 2 sampled relatively high latitudes in both hemispheres and measured a maximum field of 11.7 nT just prior to closest approach (Ness *et al.*, 1982).

More than two decades passed before the arrival of Cassini at Saturn. The Cassini spacecraft has been in orbit about Saturn since its orbit insertion (SOI) on 30 June 2004, and has remained relatively distant subsequent to the orbit insertion maneuver (Dougherty *et al.*, 2005), targeting multiple satellite flybys from the vantage point of a near-equatorial orbit. During SOI, the spacecraft periapsis was just $1.33 R_s$; this is the closest Cassini will come to Saturn throughout its mission. Cassini is instrumented with two magnetometers, a vector fluxgate and a helium magnetometer, the latter operated in a mode to measure field magnitudes near the planet (Dougherty *et al.*, 2004). However, a large number of orbits have been executed, many with periapses $<4R_s$. The spacecraft is just now executing a series of maneuvers designed to increase the inclination of the orbit, which will provide a more favorable distribution of observations for internal field analysis, particularly near the end of mission, with many orbits planned with periapses $<3R_s$.

10.07.5.1.2 Models

The Pioneer 11 magnetic field observations were consistent with a dipole field of moment $\sim 0.20 \text{ G} \cdot \text{R}^3_s$, slightly offset to the north of the planet's center (Smith *et al.*, 1980a; Acuña and Ness, 1980; Acuña *et al.*, 1980) by about 0.04 or $0.05 R_s$. The simple northward-offset

dipole model was affirmed by independent analyses of charged particle absorption signatures (Chenette and Davis, 1982) which serve as a probe of the geometry of the field. Analyses of the Pioneer 11 high-field fluxgate magnetometer observations (Acuña *et al.*, 1980) found no departure of the field from axisymmetry; however, in Saturn's relatively weak field, the resolution of the high-field fluxgate was quite limited. Working with higher-resolution VHM observations, Smith *et al.* (1980b) proposed a dipole tilt of $\sim 1^\circ$, although the orientation of the dipole could not be determined.

The combined Voyager 1 and 2 magnetic field observations were not consistent with the field of a simple displaced dipole, nor did these observations reveal any measurable departure from spin-axisymmetry (of the internal field). An axisymmetric, zonal harmonic model of degree 3 was found to be both necessary and sufficient to describe the magnetic field of Saturn (Connerney *et al.*, 1982b). This three-parameter model is referred to as the Z_3 model (zonal harmonic of degree 3) and it is listed in Table 4; the nonaxisymmetric terms ($m > 0$) are insignificant. Independent analyses of the Voyager 1 and 2 observations suggested an estimated uncertainty of $\sim 100 \text{ nT}$ for the Z_3 coefficients. The magnetic and rotation axes are indistinguishable, aligned to within $\sim 0.1^\circ$.

A series of analyses followed to test the high degree of axisymmetry of the Z_3 model (reviewed in Connerney *et al.* (1984b)). Acuña *et al.* (1983a) showed that the charged particle absorption signatures observed in Saturn's magnetosphere were consistent with the Z_3 model. A reanalysis of the absorption signatures studied by Chenette and Davis (1982), allowing for the presence of a nonzero g_3^0 term, found

Table 4 Saturn magnetic field models

	Cassini	Voyager		Pioneer 11	
Epoch	2004.5	(V1+V2)	1980.9	1981.6	1979.7
Model	SOI	Z_3	VI	V2	P11A
<i>Zonal harmonics</i>					
g_1^0	21084.	21184.	21224.	21084.	21180.
g_2^0	1544.	1606.	1678.	1607.	1980.
g_3^0	2150.	2669.	2613.	2513.	2350.
<i>External field</i>					
g_1^1		-10	-11	-8	-10
g_1^1		-1	0	-1	
h_1^1		0	0	-2	

Schmidt normalized zonal harmonic coefficients in nT, SLS coordinate system: $1R_s = 60\,330 \text{ km}$ for Voyager and Pioneer 11 models. Cassini model uses $1R_s = 60\,268 \text{ km}$. Voyager and Pioneer model coefficients provided in this table have been adjusted for a reference radius of $60\,330 \text{ km}$; see Connerney *et al.* (1982b) for coefficients referenced to $60\,000 \text{ km}$ as originally published.

these signatures to be best fit by a zonal harmonic model which was indistinguishable from the Z_3 model (Connerney *et al.*, 1984b). A reanalysis of the Pioneer 11 VHM observations (Connerney *et al.*, 1984a) identified a spacecraft roll attitude error (of 1.4° in magnitude) as the primary source of the discrepancy between the Pioneer and Voyager results. After correcting the spacecraft roll attitude, these authors demonstrated that the Pioneer 11 measurements were consistent with the Z_3 model field, and axisymmetry, to better than the accuracy of measurement (Connerney *et al.*, 1984a). The roll attitude error was subsequently independently verified by analysis of Pioneer 11 imaging photopolarimeter data. After correcting for the roll attitude error, Davis and Smith (1985, 1986) obtained a zonal harmonic model in substantial agreement with the Z_3 model (**Table 4**).

Analysis of the Cassini fluxgate magnetometer observations obtained during SOI confirmed (to the level of accuracy, approximately 0.1% and 0.1°) the spin-symmetry of Saturn's magnetic field (Dougherty *et al.*, 2005). A command error prevented the helium magnetometer from making measurements during SOI. A zonal harmonic model of degree 3 was obtained (**Table 4**) from the fluxgate observations: it is very similar to the earlier zonal harmonic models derived from Voyager and Pioneer observations, but for a smaller g_3^0 term. The Cassini mission now enters a phase in which the orbital inclination increases, providing observations of the magnetic field at progressively higher latitudes, although not as close to Saturn as SOI. With better spatial coverage, and as more observations accumulate, one may expect some refinement of the spherical harmonic models derived from Cassini.

10.07.5.1.3 Discussion

Magnetic field magnitudes (Z_3 model) on the surface of a dynamically flattened Saturn (flattening 1/10.6) range from a minimum of 0.18 G at the equator of maxima of 0.65 and 0.84 G at the southern and northern poles, respectively (**Figure 14**). Note that Saturn is the most dynamically flattened of all the planets, and the polar magnetic field is substantially increased by the decreased radial distance ($0.9 R_s$ at the poles). The equatorial field magnitude is less than that of a simple dipole or offset dipole, and polar field magnitudes (both north and south) exceed those of either a centered or offset dipole. An eccentric dipole representation of the Z_3 model has a northward displacement of the dipole of $0.038 R_s$ along the rotation axis, in agreement with the Pioneer 11 OTD representations. However, the ratio of g_3^0/g_2^0 is

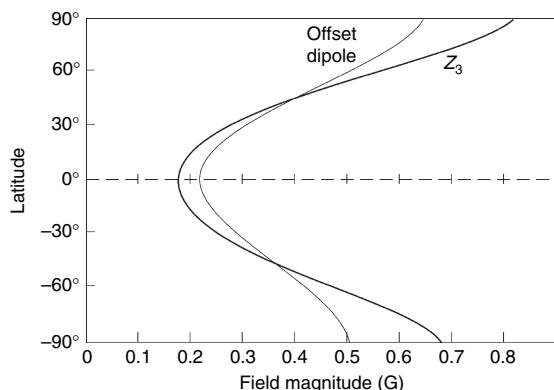


Figure 14 Magnetic field magnitude (Gauss) as a function of planetocentric latitude on the dynamically flattened (1/10.6) surface of Saturn computed using the axisymmetric Z_3 (zonal harmonic of degree 3) model of Saturn's magnetic field (Connerney *et al.*, 1982b), compared with that of a simple offset dipole.

too large by a factor of 30 to be attributed to a dipole offset. The Z_3 model's large antisymmetry with respect to the equator plane cannot be removed by a simple coordinate system translation.

The high degree of spin axisymmetry of Saturn's magnetic field is unique among the planets. This unanticipated symmetry is somewhat disquieting from an observational perspective, as it lends no simple explanation for observations of periodic magnetospheric phenomena, for example, periodicities in magnetic field and particle observations in the distant magnetosphere (Galopéau *et al.*, 1991; Connerney and Desch, 1992; Giampieri and Dougherty, 2004; Giampieri *et al.*, 2006; Cowley *et al.*, 2006; see also Kaiser *et al.*, 2005 and Giampieri *et al.*, 2005), the modulation of Saturn kilometric radio emission (Desch and Kaiser, 1981; Kaiser and Desch, 1982), optical spoke activity in Saturn's B ring (Porco and Danielson, 1982), and ultraviolet aurorae (Sandel *et al.*, 1982). All of these phenomena would presumably be easier to explain if a magnetic anomaly could be found. However, it is not clear how such a slight departure (e.g., the $\sim 0.1^\circ$ limit of detection) from axisymmetry might so dramatically influence these phenomena, which occur at different magnetic latitudes and as far as 1 or $2R_S$ above the surface.

The uncertainty regarding the origin of Saturn's quasi-periodic magnetospheric phenomena has spilled over into a related topic: Saturn's rotation period. Accurate measurement of the rotation period of gaseous planets relies on the long-term observation of magnetospheric radio emissions, which are assumed to be locked in phase with the magnetic field and interior of the planet. Long-term observation of

Jupiter's radio emissions, for example, yield a consistent estimate (Riddle and Warwick, 1976) of rotation period (9 h, 55 min, 29.71 s) regardless of frequency of emission measured or method of analysis (Carr *et al.*, 1983). Note that the uncertainty associated with this estimate is about 0.02 s (Carr *et al.*, 1983), reflecting observation over many years. Saturn's rotation period was measured in the Voyager era (1980–81) using periodicities in kilometric radio emissions (SKR), yielding an estimate of the rotation period (10 h, 39 min, 24 s) with an uncertainty of ± 7 s (Desch and Kaiser, 1981; Kaiser *et al.*, 1984). The larger uncertainty of this estimate is due to the limited span of time (9 months) over which SKR could be monitored. Continued observation of Saturn's radio emissions, however, in the Ulysses era (Galopéau and Lecacheux, 2000) and during the Cassini mission (Gurnett *et al.*, 2005; Kurth *et al.*, 2007) revealed variations of about $\pm 1\%$ (~ 6 min) in the apparent radio period – far in excess of that allowed by statistical uncertainties. Likewise, direct measurement of periodicities in the magnetic field (Giampieri and Dougherty, 2004; see also Kaiser *et al.* (2005) and Giampieri *et al.*, (2005, 2006); Cowley *et al.*, 2006) yielded inconsistent rotation periods. Saturn's radio emissions are influenced by variations in the solar wind (Desch, 1982; Desch and Rucker, 1983) and are emitted from a source region fixed in local time (Kaiser and Desch, 1982; Kaiser *et al.*, 1984; Galopéau *et al.*, 1995). It has thus been proposed that variations in the apparent periodicity of radio emissions are related to variations in the solar wind (Cecconi and Zarka, 2005). If so, it is not clear which of the proposed rotation periods most closely matches that of the deep interior of the planet. This is the subject of intense activity at present, so perhaps a better understanding of Saturn's rotation, and quasi-periodic magnetospheric phenomena, will be forthcoming soon.

The magnetic fields of Earth, Jupiter, and Mercury all favor dipole tilts of approximately 10° in magnitude; Saturn's dipole tilt is not yet determined but $< 0.1^\circ$ or so. The high degree of spin axisymmetry is surprising but not in and of itself a problem for dynamo theory (e.g., Lortz, 1972). The Cowling antidynamo theorem (Cowling, 1933, 1957), which states that regenerative dynamo action cannot occur if the magnetic field and fluid motions are axisymmetric, applies to the field in the dynamo region, not the field observed above. It is, however, necessary to explain why Saturn's magnetic field is so dramatically different from Jupiter's, given the similarity in size, composition, and rotation.

Stevenson proposed a model in which the field is axisymmetrized by the differential rotation of an

electrically conducting shell above the dynamo region (Stevenson, 1980, 1982; also Kirk and Stevenson, 1987). This nonconvecting shell forms above the dynamo region as a result of the immiscibility of helium and hydrogen under the temperature, pressure conditions in this region (the larger planet, Jupiter, would have no such region). In this model, the formation of helium raindrops results in a stably stratified conducting shell which, in differential rotation with respect to the core, attenuates any nonaxisymmetric fields generated below. Schubert *et al.* (2004) treated the more general case of dynamo generation in a convective core within an electrically conducting shell, taking into account the mutual coupling between the two. These authors conclude that the observable characteristics of planetary dynamos are largely determined by the nature of the (electrically conducting) flow above the dynamo region.

The Z_3 model of Saturn's field can be used (with caution) to predict the field inside the planet, but above the dynamo (or regions of current flow). The radial field component is of particular interest since it is continuous across the core boundary, and if the frozen flux assumption is satisfied, serves as a tracer of horizontal fluid motion at the outer radius of the dynamo (Backus, 1968; Benton, 1979; Benton and Muth, 1979). Figure 15 shows the latitudinal variation of $|B_r|$, computed using the Z_3 model, on the surface of a sphere of radius $r = 0.5 R_s$, which is the approximate outer boundary of the metallically conducting core (Stevenson, 1983). The Z_3 model (via the g_3^0 term) enhances the field at the poles and reduces the radial field near the equator, relative to a simple dipole. In fact, the Z_3 model is very nearly identical to that required to minimize the low-latitude unsigned flux across the core surface (Connerney, 1993). This may be the signature of a dynamo axisymmetrized by differential rotation of a conductive layer above.

The high degree of symmetry (about the rotation axis) of Saturn's magnetic field is a distinguishing feature that may have some unexpected consequences throughout the system. For this planet, a specific radial distance in the equatorial plane maps (along the magnetic field) uniquely to a corresponding latitude on the surface (Connerney, 1986), as is illustrated in Figure 16. Thus, one may associate a radial distance (ring plane conjugate) with any northern or southern latitude. Narrow dark bands that encircle the planet at 44.2° N, 46.3° N, and 65.5° N have been attributed (Connerney, 1986) to an influx of water guided along magnetic field lines from sources in the ring plane: two locations at the inner edge of the B ring ($1.525 R_s$, $1.62 R_s$) where high

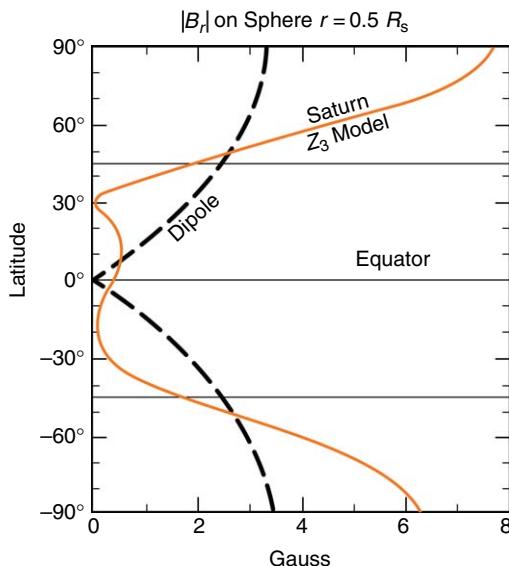


Figure 15 Unsigned magnetic flux computed on a sphere of radius $0.5R_s$ (presumed dynamo core radius) using the Z_3 magnetic field model, compared to that for a simple dipole. Saturn's magnetic field is very nearly that which (with three terms) minimizes the low-latitude ($<45^\circ$) unsigned flux across the core boundary. From Connerney JEP (1993) Magnetic fields of the outer planets. *Journal of Geophysical Research* 98(E10): 18659–18679.

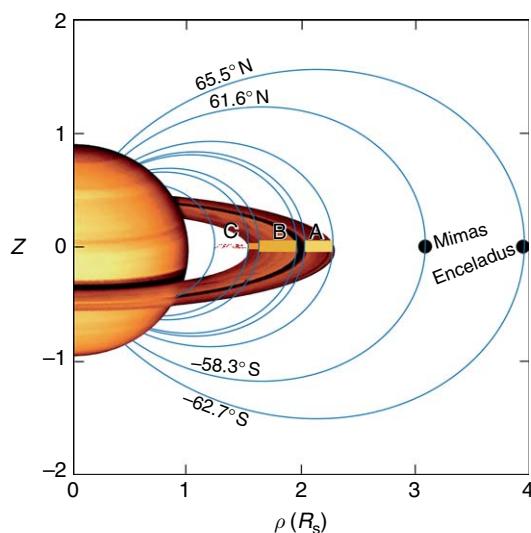


Figure 16 Saturn's magnetic field is very nearly symmetric about its rotation axis but not about the equator; all field lines passing through the ionosphere at one latitude cross the ring plane at a unique radial distance (ring plane conjugate) and map to a corresponding (but different) latitude in the opposite hemisphere. The magnetic equator is at $5.9^\circ N$ (planetocentric) latitude (Connerney, 1986).

charge-to-mass ratio particles (e.g., submicron ice or dust) become unstable (Northrop and Connerney, 1987) and a third ($3.95 R_s$) at the orbit of Enceladus and the E ring. The B-ring sources are identified with and electromagnetic erosion mechanism that sculpted the rings (Northrop and Connerney, 1987). The Enceladus/E-ring source may reflect persistent activity of the recently discovered geysers on Enceladus (Kivelson, 2006; Kargel, 2006).

10.07.6 Ice Giants

10.07.6.1 Uranus

Uranus, residing at a mean orbital distance of 19.2 AU, was a mystery prior to the Voyager 2 encounter with Uranus in January 1986. The planet appears bland and featureless when observed from the Earth even under the best circumstances. Uranus and Neptune are similar in size (approximately 25 000 km radius), mass (about 15 Earth masses), and presumably composition (largely low-temperature condensates, or ‘ice’: H_2O , NH_3 , CH_4). One might therefore expect that their magnetic fields would be similar. Uranus is unique in that its rotation axis lies very nearly in its orbital plane, which is within 1° of the ecliptic plane. The actual inclination of its equatorial plane to that of its orbit is 82° . Twice during each orbit about the Sun (84 years), the rotation axis is oriented very nearly along the planet–Sun line, so that we on Earth look upon the pole. This was the case during the Voyager 2 encounter in 1986, the Southern Hemisphere was directed toward the Sun.

Uranus was not known to have a magnetic field prior to the Voyager encounter, having maintained ‘radio silence’ prior to encounter. It is a relatively weak radio source, emitting only at lower frequencies (below about 1 MHz), making detection from Earth difficult. In addition, the peculiar ‘pole-on’ encounter geometry kept the Voyager spacecraft beyond the reach of radio emissions. Voyager 2, on a flyby trajectory (Figure 17) approaching to within $4.2R_u$, provided the first and thus far only measurements of the magnetic field (Ness et al., 1986). These measurements led to the rather surprising discovery of the first ‘oblique rotator’, a planet with a large angular separation between magnetic and rotation axes.

10.07.6.1.1 Models

Preliminary analyses of the Voyager magnetic field observations used an OTD representation of the field. This model is characterized by a dipole of moment

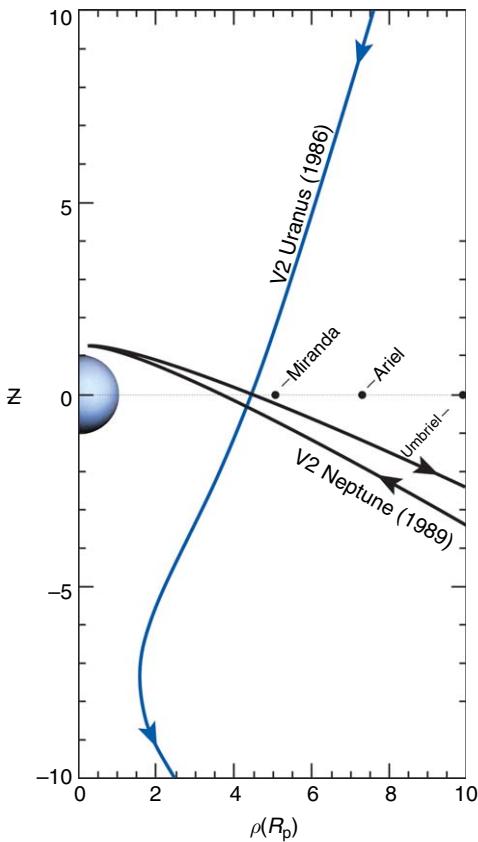


Figure 17 Voyager 2 spacecraft flybys of Uranus and Neptune in a cylindrical coordinate system. The spacecraft distance from the equator plane (z) as a function of distance from the rotation axis (ρ) is given in units of planet radius. The positions of satellites Miranda, Ariel, and Umbriel in the Uranus system are indicated.

0.23 G R^3_u ($IR_u=25\,600$ km), displaced along the rotation axis by $0.33R_u$ and inclined by 60° with respect to the rotation axis (Ness *et al.*, 1986). Since Voyager remained somewhat distant, and OTD representation is expected to be a good approximation to the field, and indeed it fits the observations well (Ness *et al.*, 1986). The very large dipole offset of the OTD is an indication of a planetary magnetic field with a large quadrupole, and likely higher-degree and-order terms, that is, a very complex field.

A spherical harmonic analysis of the field required an expansion to degree and order 3 (octupole) to adequately represent the measurements (Connerney *et al.*, 1987), which is somewhat surprising in view of Voyager's relatively large periapsis. The Voyager trajectory provided a fairly good spatial distribution of observations, sufficient to resolve, or determine, spherical harmonic terms of degree 1 and 2, the dipole and quadrupole. This partial solution to the underdetermined inverse problem, designated 'Q₃'

Table 5 Uranus and Neptune magnetic field models

Model	Uranus ($1R_u = 25\,600$ km)	Neptune ($1R_n = 24\,765$ km)
1	g_1^0 11893.	Neptune O ₈ 9732.
2	g_1^1 11579.	3220.
3	h_1^1 -15684.	-9889.
4	g_2^0 -6030.	7448.
5	g_2^1 -12587.	664.
6	g_2^2 196.	4499.
7	h_2^1 6116.	11230.
8	h_2^2 4759.	-70.
9	g_3^0 *	-6592. *
10	g_3^1 *	4098. *
11	g_3^2 *	-3581. *
12	g_3^3 *	484.
13	h_3^1 *	-3669.
14	h_3^2 *	1791. *
15	h_3^3 *	-770.

Uranus model from Connerney *et al.* (1987); see reference for parameter uncertainties and resolution. Neptune model from Connerney *et al.*, 1991; see this reference for discussion of parameter resolution.

(Quadrupole part of an expansion to degree and order 3), is listed in **Table 5**. The relative magnitude of the quadrupole is quite large; the octupole is likely to be large as well, but cannot be meaningfully constrained by the observations. Lack of knowledge of higher-degree and-order terms may be an important consideration, particularly in estimation of the field near the planet. The Q₃ model provides a low-degree and-order approximation only, and it is likely that surface fields are more complex than illustrated here.

The magnetic field magnitude on the surface of Uranus reaches a maximum of approximately 1 G at mid-latitudes in the south, and a minimum of ≤ 0.1 G at middle to high northern latitudes (**Figure 18**). **Figure 18** also illustrates the foot of the Miranda flux tube, which is the path traced out on the planet's surface by the field lines passing through the satellite as the planet rotates. Also shown are estimates of the position of the aurorae (stippled regions) and the paths of multiple-dip equators (where $B_r=0$). Uncertainties in the higher-degree and-order components of the field, which are substantial, may lead to large differences in the actual surface field, particularly in the weaker field regions. Note that for Uranus, the convention adopted by the Voyager particles and fields investigations is at odds with the International Astronomical Union (IAU) latitude convention. The Voyager system has planetary latitudes positive over the hemisphere in which the angular momentum vector resides, that is, the

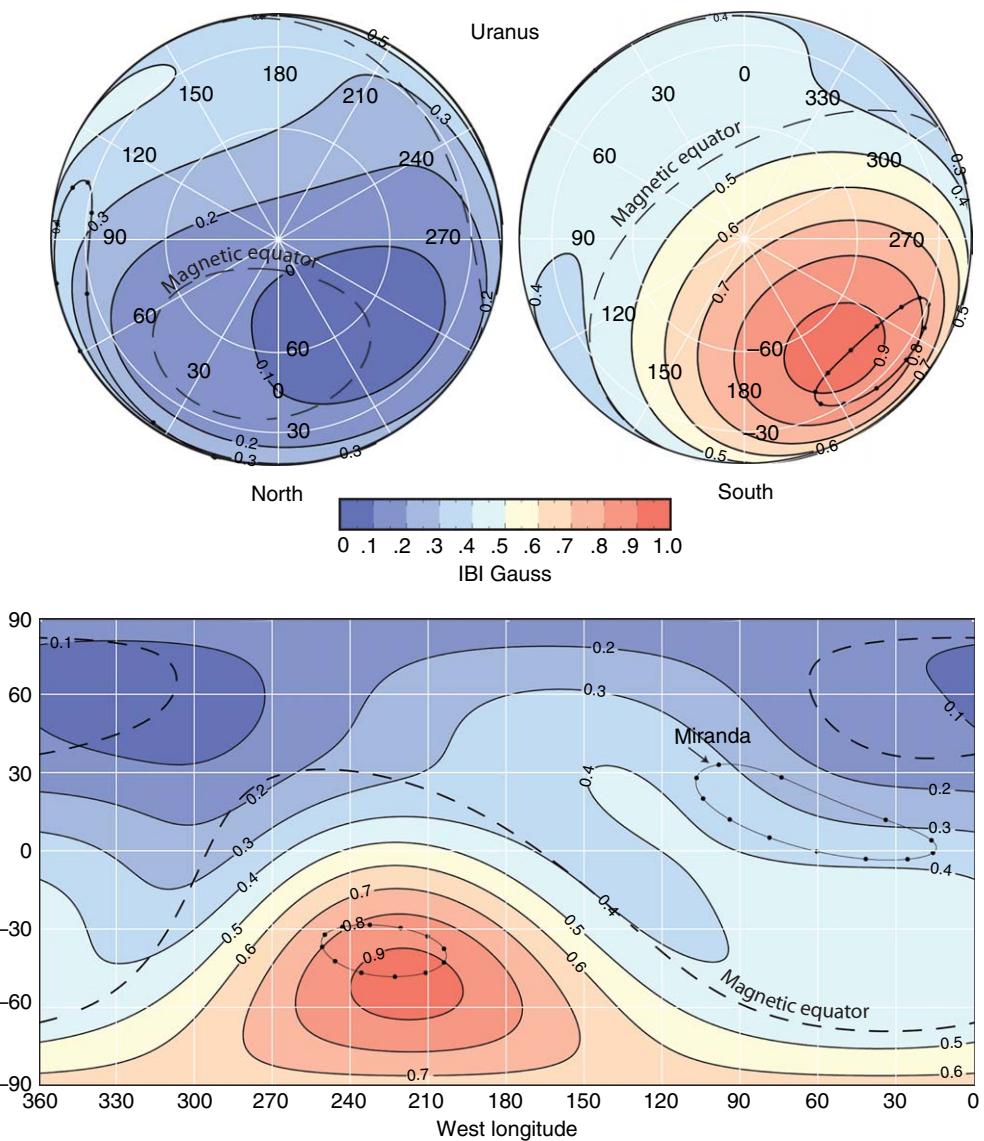


Figure 18 Contours of constant magnetic field magnitude (Gauss) on the dynamically flattened (1/41.6) surface of Uranus, computed using the Q_3 spherical harmonic model (see text). The top panel shows the field magnitude in orthographic projections viewed from $+z$ (left) and $-z$ (right); below the color bar is a rectangular latitude-longitude (west longitude) projection. A trace (dashed) indicates the position of the magnetic equator. Solid lines join points on the surface that trace along field lines to the orbit of the satellite Miranda, with filled circles for increments of 30° in the satellite's longitude.

hemisphere oriented toward the Sun (and Earth) at the time of the Voyager encounter. The IAU defines the south pole as the rotation pole south of the ecliptic, without regard to the direction of rotation.

10.07.6.2 Neptune

Neptune, residing at a mean orbital distance of 30 AU, is the most distant of the Jovian planets. It is similar in size (approximately 25 000 km radius), mass (about 15 Earth masses), and composition (largely low-

temperature condensates, or ‘ice’: H_2O , NH_3 , CH_4) to Uranus. Very little was known of Neptune prior to the Voyager 2 encounter in August 1989. The first evidence of Neptune’s magnetic field appeared in radio observations (Warwick *et al.*, 1989) obtained by Voyager 2 just days prior to its closest approach to the planet on 25 August 1989.

The *in situ* magnetic field observations revealed a magnetic field of internal origin with characteristics not unlike those of the planet Uranus (Ness *et al.*, 1989). The most striking difference between the Uranus and

Neptune encounter was a difference of our own design: the Voyager 2 flyby trajectories at each planet. Where Voyager 2 at Uranus remained relatively distant from the planet, Voyager 2 at Neptune approached to within $1.18R_n$ of the planet (Figure 17). The presence of higher-degree and-order components of the field in the latter observations is overwhelming, simply a consequence of the close periapsis. Preliminary analyses of the Voyager magnetic field observations used an OTD representation of the field, valid at moderate distances from the planet ($4 - 15R_n$ radial distance; $1R_n = 24\,765$ km), and fitted to data in excess of $4R_u$ from the planet only. In this first approximation, the field may be characterized as that of a dipole of moment $0.13\text{ G }R_n$, offset from the center of the planet by a rather surprising large $0.55R_n$ ³ and inclined by 47° with respect to the rotation axis (Ness *et al.*, 1989).

It was necessary for Ness *et al.* (1989) to exclude from their OTD analysis all observations of the field obtained at $r \leq 4R_n$. One indication of how nondipolar the field of Neptune is can be obtained by comparison of the actual field measured at closest approach (approximately 10 000 nT) to that one would expect if the field were dipolar (6500 nT, for the OTD model). In fact, the field measured near closest approach was quite complex, as is illustrated in Figure 19. Two local maxima appear in the magnitude of the field observed along the flyby

trajectory, instead of the one maximum one would observe in a dipolar field. Spherical harmonic analysis of the field required an expansion to degree and order 8 to adequately represent the measurements (Connerney *et al.*, 1991). The large number of parameters associated with a spherical harmonic expansion of degree and order 8 (80 Schmidt coefficients) and the rather limited observations leads to a severely under-determined inverse problem which has only a partial solution (Connerney *et al.*, 1991). The distribution of observations is sufficient to resolve, or determine, all of the spherical harmonic terms of degree 1 (dipole), some of the terms of degree 2 (quadrupole), and some terms of degree 3 (octupole) and higher. The resulting model, designated ‘ O_8 ’ (octupole part of an expansion to degree and order 8), is listed in Table 5. For the Neptune field model, it is necessary to consider the resolution of each parameter listed in Table 5. For this we use the magnitude of the corresponding diagonal element of the resolution matrix, a measure of the uniqueness of the solution, the extent to which an individual model parameter can be represented by the basis vectors included in the partial solution (see Connerney *et al.*, 1991). Note that some of the quadrupole and octupole coefficients of the O_8 model are not well resolved, but included out of necessity (see discussion of parameter resolution in Connerney *et al.* (1991)).

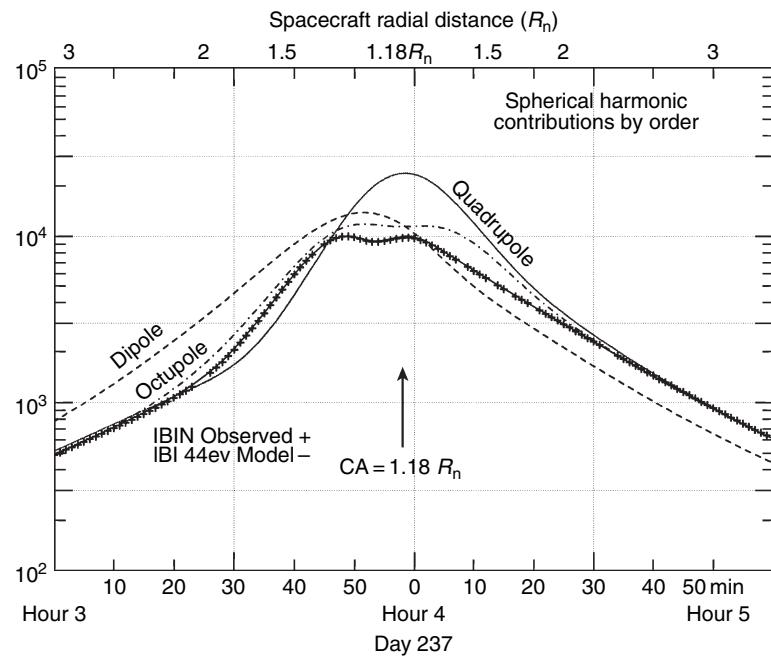


Figure 19 Magnitude of the observed magnetic field (crosses) as a function of time (lower axis) and spacecraft radial distance (upper axis) for the Voyager 2 encounter with Neptune, compared with a model fit (O_8 model) and the relative magnitudes of the dipole (dashed), quadrupole (solid), and octupole (dot-dashed) terms in the spherical harmonic expansion.

As anticipated, the relative magnitude of the quadrupole is quite large; the octupole is large as well, but can only be loosely constrained by the Voyager observations.

The magnetic field magnitude on the surface of the planet ranges from approximately 0.9 G at southern mid-latitudes to ≤ 0.1 G near the equator and at northern mid-latitudes (**Figure 20**). **Figure 20** also shows estimated auroral zones (stippled regions) which differ greatly in size as a result of the hemispherical asymmetry in field magnitudes. In this representation, it must

be recognized that only the low-degree and -order part of the field is portrayed on the surface of the planet, and the likely presence of higher-degree and -order terms may be expected to significantly alter the surface field. It is clear that the magnetic fields of Uranus and Neptune are quite similar in the range of surface field magnitudes, the tilt of the dipole, and the appearance of one strong pole at southern mid-latitudes. Both Uranus and Neptune have extended weak-field regions, multiple-dip equators, and multipolar structure in evidence at the surface of the planet.

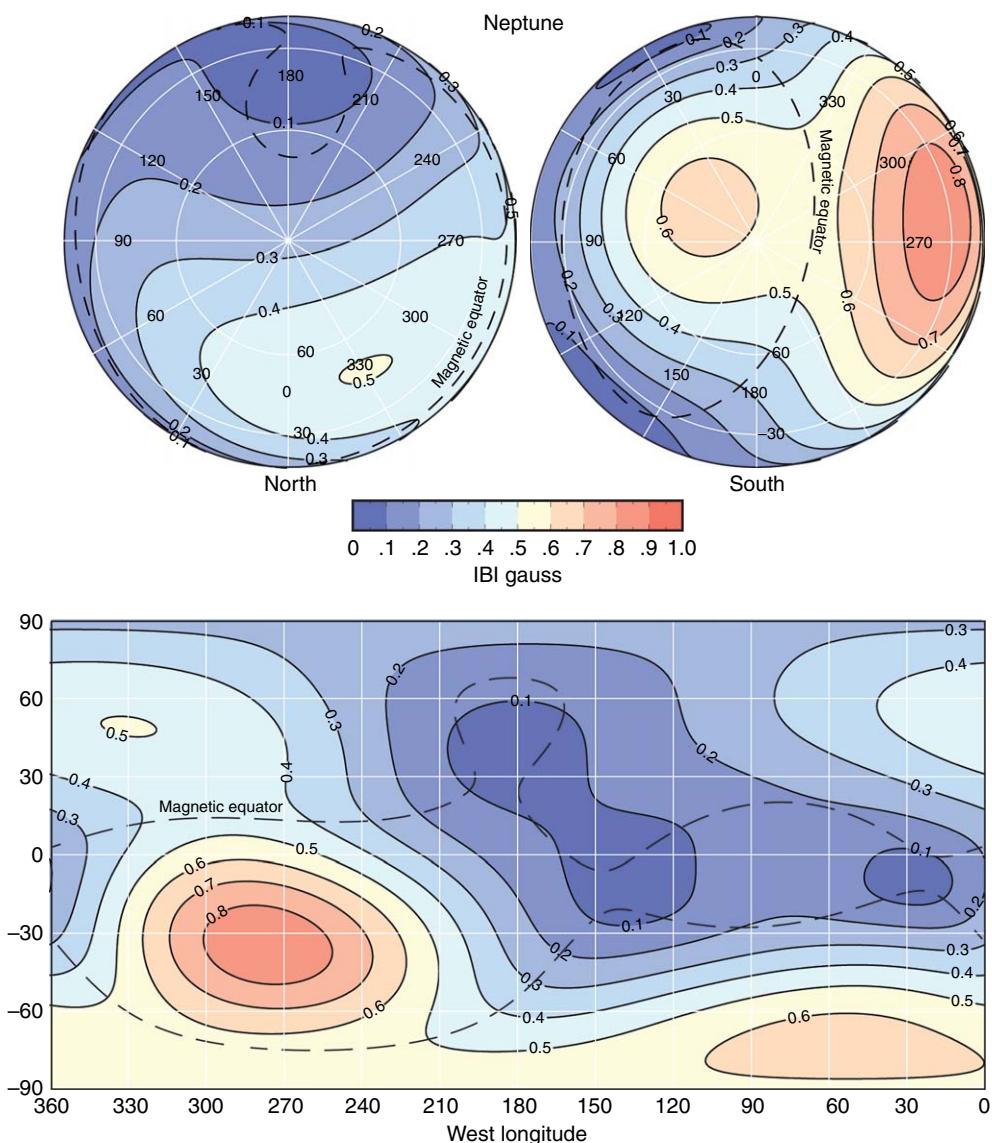


Figure 20 Contours of constant magnetic field magnitude (G) on the dynamically flattened (1/47.6) surface of Neptune, computed using the O_8 spherical harmonic model (see text). The top panel shows the field magnitude in orthographic projections viewed from $+z$ (left) and $-z$ (right); below the color bar is a rectangular latitude-longitude (west longitude) projection. A trace (dashed) indicates the position of the magnetic equator.

10.07.7 Satellites and Small Bodies

10.07.7.1 Moon

The Moon, with a radius of 1738 km and a mass of little more than 1% that of the Earth ($m/M_e = 0.0123$), is the largest satellite among the terrestrial planets. It is not so much smaller than Mercury (2439 km) in size, with a little more than one-third the volume of Mercury, but only about one-fifth of the mass. It is a differentiated object similar in composition to the Earth's mantle.

The Moon has been a frequent target of space exploration spanning nearly half a century, beginning with a series of robotic probes in the 1960s (Ranger series of the US space program, Luna and Zond series of the USSR program) and culminating in the first missions of human exploration of the Moon, the US Apollo Program. For a brief moment in history, beginning with Apollo 11 on 20 July 1969, and ending with Apollo 17 on 14 December 1972, humans walked (and drove) on the surface of another world, deploying instruments and collecting soil and rock samples for return to Earth. Much of what we know of lunar

magnetism evolved from study of the lunar samples returned from the Moon, combined with *in situ* observations collected both on the surface and in orbit about the Moon. Robotic scientific exploration of the Moon was rekindled decades after the Apollo Program ended, beginning with the orbiting Clementine spacecraft in February and March of 1994. The Lunar Prospector spacecraft was in polar mapping orbit from 13 January 1998 through 31 July 1999, mapping crustal magnetic fields via both direct and indirect methods. ESA's SMART-1 spacecraft was in orbit about the Moon from 15 November 2004 through impact on 3 September 2006.

The most complete Lunar magnetic field map (Figure 21) currently available was compiled using 1.5×10^5 estimates of the surface field magnitude obtained over Lunar Prospector's 18-month mission (Mitchell *et al.*, 2007). This recent map uses the electron reflection technique corrected for charging of the lunar surface and spacecraft (Halekas *et al.*, 2002), significantly extending (to ~ 0.2 nT) the dynamic range of field magnitudes mapped relative to previous maps

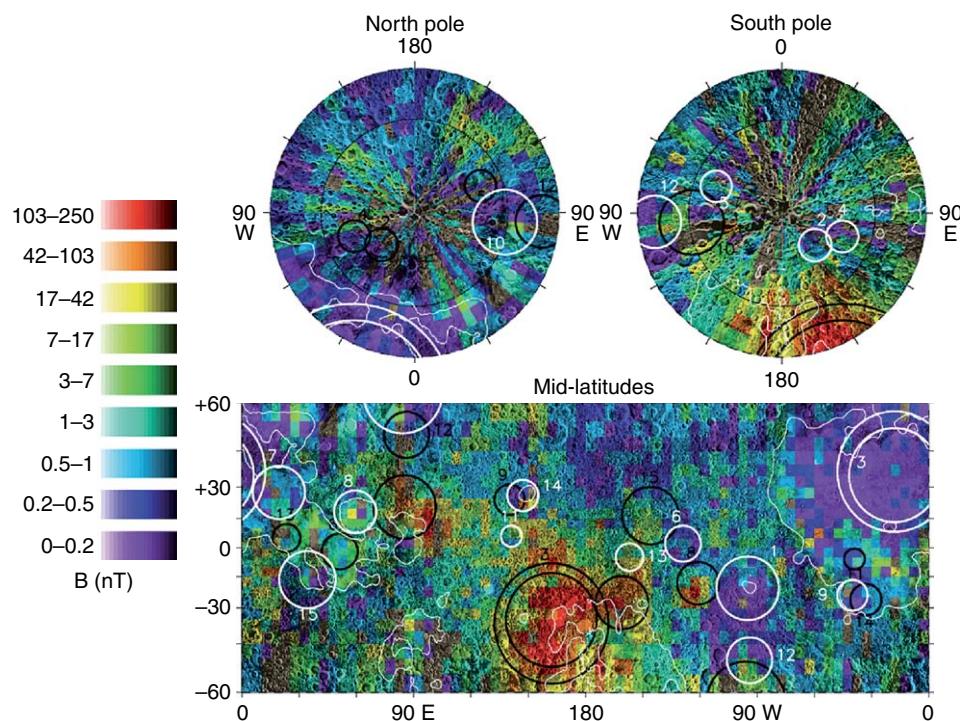


Figure 21 Magnetic field magnitude on the Lunar surface as measured by the electron reflection technique. Field magnitudes from 0.2 to 250 nT are represented by colors according to the color bar to the left. The top panel shows the field magnitude in orthographic projections viewed from $+z$ (left) and $-z$ (right); below is a rectangular latitude-longitude projection. The white circles denote major impact basins, and the black circles are drawn at the antipodal site. From Mitchell DL, Halekas JS, Lin RP, Frey S, Hood LL, Acuña MH, and Binder A (2007) A global mapping of Lunar crustal fields by Lunar prospector. *Journal of Geophysical Research* (in press).

(Halekas *et al.*, 2001). This map is still undersampled, in the sense that substantially more observations would be required to obtain complete spatial coverage at the resolution implied by the electron gyroradius (about 5 km at 300 eV). However, it provides a good sense of the global distribution of magnetic sources in the Lunar crust and the characteristic scale length of coherent magnetization in the Lunar crust. Where the Lunar crust is magnetized, it is weakly magnetized (on spatial scales of 5 km and greater), compared to the Mars crust, or Earth's. Lunar crustal magnetization also appears to be less coherent at large spatial scales, compared to Earth and Mars. The large impact basins (e.g., Orientale, Imbrium) appear nonmagnetic, implying that the impact melt cooled in zero field (no lunar dynamo at the time of impact); and yet, the regions of strongest field magnitude appear 'antipodal' to the largest impact basins (Lin *et al.*, 1988). So it seems that Lunar crustal magnetization and large impacts are related; perhaps impacts magnetize the crust not at the impact site, but on the opposite side of the Moon. One theory for how this might occur (Hood and Vickery, 1984; Hood and Huang, 1991) uses the impact-generated plasma cloud to intensify the weak ambient magnetic field at the antipode, if only briefly (~ 1 day), as it expands from the impact site to envelop the Moon. The expanding plasma cloud, being a good electrical conductor, sweeps up the ambient field as it converges on the antipode, where the crust then acquires remanence via shock magnetization (Fuller *et al.*, 1974; Cisowski *et al.*, 1983).

10.07.7.2 Ganymede

Ganymede, with a radius of 2634 km, is the largest satellite in the solar system. It is slightly larger than the smallest planet in our solar system, Mercury ($R_m = 2439$ km), but only about half as massive ($M_g = 0.025 M_e$). Ganymede may have an Fe or FeS core extending to as much as half the satellite's radius, overlain by a silicate mantle and topped with ~ 800 km layer of ice (Anderson *et al.*, 1996).

The Galileo spacecraft was targeted for several close flybys of Ganymede, the first of which occurred in June of 1996. The first two flyby encounters, with minimum altitudes of 838 and 264 km, demonstrated conclusively the presence of an intrinsic dipolar magnetic field and associated magnetosphere (Kivelson *et al.*, 1996, 1997; Gurnett *et al.*, 1996). Initial analysis of the magnetometer observations revealed a dipole field with equatorial surface field of 750 nT, nearly (within 10°) aligned with the rotation axis. A field of

this magnitude is sufficient to produce a Ganymede 'mini-magnetosphere' with all of the characteristics normally associated with a trapped particle population. Ganymede's immediate environment (Figure 22) includes field lines with both ends anchored within Ganymede, hosting a local trapped radiation environment; 'open' field lines that extend to the Jovian ionosphere; and Jovian field lines. A more comprehensive analysis, performed after the discovery of induced magnetic fields associated with Europa and Callisto (Neubauer, 1998; Khurana *et al.*, 1998; Kivelson *et al.*, 1999; Zimmer *et al.*, 2000), allowed for the presence of an induced field in addition to the main field (Kivelson *et al.*, 2002). This analysis demonstrated that Ganymede has both an inductive response to the time-varying magnetic field (due to currents induced in a hidden global ocean) and an intrinsic field likely due to a dynamo in the deep interior. The intrinsic field is a permanent dipole with a surface equatorial field of 719 nT, tilted by 176° with respect to the rotation axis and rotated 24° toward the trailing hemisphere from the Jupiter-facing meridian plane. The permanent field is perhaps the best evidence that Ganymede has a substantial Fe or FeS core in the interior; dynamo action implies that it is fluid, and

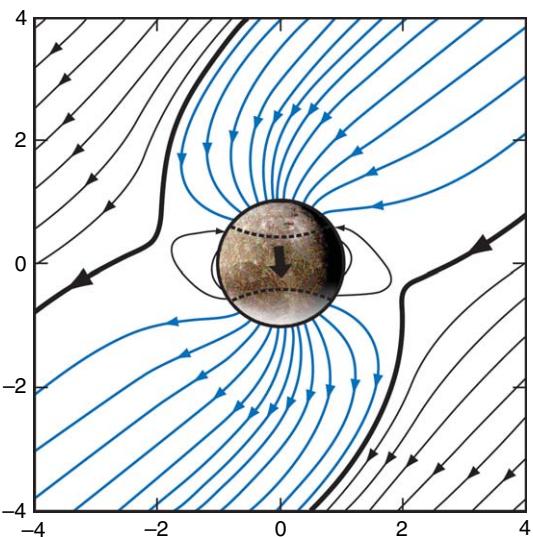


Figure 22 Magnetic field of Ganymede and its immediate environment in meridian plane projection. Ganymede's internal magnetic field moment is indicated with a filled arrow. Low-latitude field lines close within the satellite's interior; high-latitude field lines (blue) form a flux tube with one end rooted in Jupiter's ionosphere; Jovian field lines (black) are separated from those linking Ganymede by a separatrix (heavy black). Adapted from Kivelson MG, Khurana K, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.

convecting, with sufficient energy input to sustain the dynamo. However, it is possible (but not likely) that Ganymede's dipolar magnetic field is due to remanent magnetism (Crary and Bagenal, 1998).

10.07.8 Discussion

A convenient measure of the complexity of a planetary magnetic field, often used in studies of the magnetic field of the earth and planets, is the 'harmonic spectrum,' sometimes referred to as a 'Lowes spectrum,' defined as follows (Lowes, 1974; Langel and Estes, 1982):

$$R_n = (n+1) \sum_{m=0}^n \left\{ (g_n^m)^2 + (h_n^m)^2 \right\}$$

This quantity is equal to the mean-squared magnetic field intensity over the planet's surface produced by harmonics of degree n . Scaled to the core–mantle boundary with the factor $(a/r_c)^{2n+4}$, this quantity represents the mean-squared magnetic field intensity at the dynamo surface. The Earth's field is well known to high degree and order ($N_{\max} = 23$). Scaled to the core–mantle boundary, the spectrum becomes almost flat for $n \leq 14$, suggesting a 'white' spectrum for the dynamo at the core–mantle boundary (e.g., Lowes, 1974). It is assumed that the core–mantle boundary, the location of which is very accurately known, represents the outer boundary of the geodynamo. Thus, it has often been assumed that a white spectrum is a common feature of all planetary dynamos (Elphic and Russell, 1978), although in the Earth's case the quadrupole is considerably less than expected. In Figure 23, R_n is calculated using the GSFC 12/83 model for the Earth (Langel and Estes, 1985), the Z_3 model for Saturn (Connerney *et al.*, 1982b), and the Q_3 model for the magnetic field of Uranus (Connerney *et al.*, 1987). For Jupiter both the O_4 model (Acuña and Ness, 1976) and O_6 models (Connerney, 1992) are shown, illustrating a range of possible values for the quadrupole and octupole. For Uranus, no estimate of the magnitude of higher-degree ($n > 2$) moments can be made, while for Neptune, terms of higher degree are appreciable, but quite uncertain.

For Jupiter and Saturn, we assume that the core radius is near the radius of the metallically conducting hydrogen core, approximately 0.75 and 0.5 planet radius, respectively (e.g., Stevenson, 1983). For Uranus and Neptune, we take $r_c = 0.75 R_p$ with the expectation that the dynamo for these planets operates in the ice mantle, where ionic conduction prevails. In

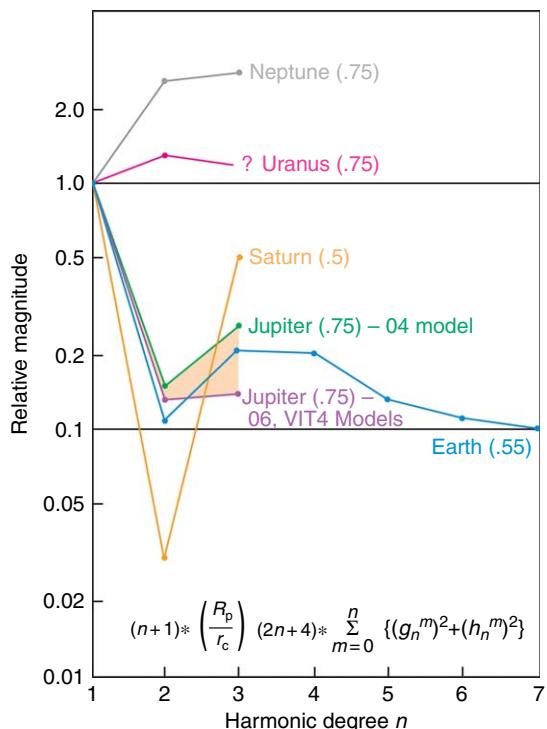


Figure 23 Relative harmonic content (Lowe's spectrum) of spherical harmonic models of Jupiter, Saturn, Uranus, and Neptune, compared with that of Earth, normalized to the assumed core radius (in parentheses) for each planet. Two Jovian field models are chosen to illustrate the range of possibilities for that planet.

this presentation, the planets fall into two distinct groups, one in which the core surface field is dominated by the dipole (Earth, Jupiter, Saturn) and one in which the core surface field is dominated by higher-degree contributions (Uranus and Neptune). However, in recognition of the peculiarity of Saturn's axisymmetric field, one might better think in terms of three classifications. In Stevenson's (1982) model, differential rotation of an outer conductive shell attenuates the nonaxisymmetric components of the dynamo within, thereby reducing the observed tilt by an order of magnitude or more from the 'nominal' value. This model has an appealing physical basis (material properties of the H–He fluid) and appears to satisfy other constraints, for example, helium depletion of the gaseous outer envelope and a relatively high planetary heat flux. The dipole-dominated group includes the planets with near alignment of dipole and rotation axes ($\leq 10^\circ$); the other includes the oblique rotators Uranus (59°) and Neptune (47°), the planets with a very large angular separation of the dipole and rotation axes.

10.07.9 Summary

The magnetic planets (and satellites) of the solar system provide a rich diversity of objects of relevance to dynamo theory. Mercury, Mars, Earth, and possibly Ganymede, have substantial iron cores with varying amounts of S or some other alloying element; each has, or once had, an active dynamo. In this group Mars and Earth appear similar, although the Mars dynamo is no longer active, and we have only the crustal remanence that remains today as evidence of its dynamo. Mercury and Ganymede are similar in size as well as magnetic properties; both have diminutive dynamos often considered so weak that alternatives to an active dynamo are proposed (Giampieri and Balogh, 2002; Aharonson *et al.*, 2004; Crary and Bagenal, 1998; however, see Christensen (2006)). Both have dipolar, or nearly dipolar fields more or less aligned with the spin axis.

Unmagnetized Venus reminds us that in dynamo theory, the ‘devil is in the details’.

In Jupiter and Saturn, we have two planets of nearly the same composition and comparable size but with magnetic fields that are worlds apart. Jupiter’s dynamo produces a magnetic field resembling Earth’s, while Saturn’s is uniquely spin-axisymmetric. Stevenson’s account of Saturn’s internal structure and magnetic field is appealing, as it provides a vehicle for understanding the differences between these giant planets. In Uranus and Neptune we have again two planets of similar composition and size, this time with similar magnetic fields (fortunately!). These icy twins form a category of ‘oblique rotators’ with large dipole tilts and eccentric magnetic fields. Heuristically, it is tempting to associate dipole dominance with small dipole tilt and dipole deficiency with large dipole tilt; it appears that dynamos in less conductive fluids are in the latter group.

References

- Acuña MH and Ness NF (1976) Results from the GSFC fluxgate magnetometer on Pioneer 11. In: Gehrels T (ed.) *Jupiter*, pp. 830–847. Tucson, AZ: University of Arizona Press.
- Acuña MH and Ness NF (1980) The magnetic field of Saturn: Pioneer 11 observations. *Science* 207: 444–446.
- Acuña MH, Ness NF, and Connerney JEP (1980) The magnetic field of Saturn: Further studies of the Pioneer 11 observations. *Journal of Geophysical Research* 85: 5675–5678.
- Acuña MH, Connerney JEP, and Ness NF (1983a) The Z_3 zonal harmonic model of Saturn’s magnetic field: Analyses and implications. *Journal of Geophysical Research* 88: 8771–8778.
- Acuña MH, Behannon KW, and Connerney JEP (1983b) Jupiter’s magnetic field and magnetosphere. In: Dessler AJ (ed.) *Physics of the Jovian Magnetosphere*, pp. 1–50. New York: Cambridge University Press.
- Acuña MH, Connerney JEP, Wasilewski P, *et al.* (1992) Mars Observer magnetic fields investigation. *Journal of Geophysical Research* 97: 7799–7814.
- Acuña MH, Connerney JEP, Wasilewski P, *et al.* (1998) Magnetic field and plasma observations at Mars: Initial results of the Mars Global Surveyor Mission. *Science* 279: 1676–1680.
- Acuña MH, Connerney JEP, Ness NF, *et al.* (1999) Global distribution of crustal magnetism discovered by the Mars Global Surveyor MAG/ER Experiment. *Science* 284: 790–793.
- Acuña MH, Connerney JEP, Wasilewski P, *et al.* (2001) Magnetic field of Mars: Summary of results from the aerobraking and mapping orbits. *Journal of Geophysical Research* 106: 23403–23417.
- Aharonson O, Zuber MT, and Solomon SC (2004) Crustal remanence in an internally magnetized non-uniform shell: A possible source for Mercury’s magnetic field? *Earth and Planetary Science Letters* 218: 261–268.
- Albee AL, Palluconi FD, and Arvidson RE (1998) Mars global surveyor mission: Overview and status. *Science* 279(5357): 1671–1672.
- Anderson JD, Lau EL, Sjogren WL, Schubert G, and Moore WB (1996) Gravitational constraints on the internal structure of Ganymede. *Nature* 384(6609): 541–543.
- Anderson KA, Lin RP, McGuire RE, and McCoy JE (1975) Measurement of lunar and planetary magnetic fields by reflection of low energy electrons. *Space Science Instruments* 1: 439–470.
- Anselmi A and Scoon GEN (2001) BepiColombo, ESA’s Mercury Cornerstone mission. *Planetary Space Science* 49(14–15): 1409–1420.
- Arkani-Hamed J (2001a) A 50-degree spherical harmonic model of the magnetic field of Mars. *Journal of Geophysical Research* 106(E10):23: 23197–23208.
- Arkani-Hamed J (2001b) Paleomagnetic pole positions and pole reversals of Mars. *Geophysical Research Letters* 28(17): 3409–3412.
- Arkani-Hamed J (2002) An improved 50-degree spherical harmonic model of the magnetic field of Mars derived from both high-altitude and low-altitude data. *Journal of Geophysical Research* 107(E5) (doi:10.1029/2001JE001496).
- Backus GE (1968) Kinematics of secular variation in a perfectly conducting core. *Philosophical Transactions of the Royal Society of London* 263: 239–266.
- Backus GE (1986) Poloidal and toroidal fields in Geomagnetic field modeling. *Reviews of Geophysics* 24: 75–109.
- Balogh A, Dougherty MK, Forsyth RJ, *et al.* (1992) Magnetic field measurements in the vicinity of Jupiter during the Ulysses flyby. *Science* 257: 1515–1518.
- Bergen S and Engle IM (1981) Mercury magnetosphere and magnetotail revisited. *Journal of Geophysical Research-Space Physics* 86(NA3): 1617–1620.
- Benton ER (1979) Magnetic probing of planetary interiors. *Physics of the Earth and Planetary Interiors* 20: 111–118.
- Benton ER and Muth LA (1979) On the strength of electric currents and zonal magnetic fields at the top of the Earth’s core: Methodology and preliminary estimates. *Physics of the Earth and Planetary Interiors* 20: 127–133.
- Berge GL and Gulkis S (1976) Earth-based radio observations of Jupiter: Millimeter to meter wavelengths. In: Gehrels T (ed.) *Jupiter*, pp. 621–692. Tucson, AZ: University of Arizona Press.
- Bigg EK (1964) Influence of the satellite Io on Jupiter’s decametric emission. *Nature* 203: 1008–1010.
- Breuer D and Spohn T (2003) Early plate tectonics versus single plate tectonics on Mars: Evidence from magnetic field

- history and crust evolution. *Journal of Geophysical Research* 108(E7): 5072.
- Burke BF and Franklin KL (1955) Observations of a variable radio source associated with the planet Jupiter. *Journal of Geophysical Research* 60: 213–217.
- Cain JC, Ferguson B, and Mozzoni D (2003) An $n = 90$ internal potential function of the magnetic field of the Martian crustal magnetic field. *Journal of Geophysical Research* 108(E2): 5008 (doi:10.1029/2000JE001487).
- Carr TD, Desch MD, and Alexander JK (1983) Phenomenology of magnetospheric radio emissions. In: Dessler AJ (ed.) *Physics of the Jovian Magnetosphere*, pp. 266–284. New York: Cambridge University Press.
- Caudal G and Connerney JEP (1989) Plasma pressure in the environment of Jupiter, inferred from Voyager 1 magnetometer observations. *Journal of Geophysical Research* 94: 15055–15061.
- Cecconi B and Zarka P (2005) Model of a variable radio period for Saturn. *Journal of Geophysical Research* 110: A12203 (doi:10.1029/2005JA011085).
- Chapman S and Bartels J (1940) *Geomagnetism*, pp. 639–668. New York: Oxford University Press.
- Chenet D and Davis L, Jr. (1982) An analysis of the structure of Saturn's magnetic field using charged particle absorption signatures. *Journal of Geophysical Research* 87: 5267–5274.
- Chicarro A, Martin P, and Trautner R (2004) The Mars Express mission: An overview. In: Wilson A (ed.) *ESA Special Publication-1240: Mars Express: An European Mission to the Red Planet*, pp. 3–13. Noordwijk, The Netherlands: ESTEC.
- Christensen U (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 444: 1056–1058 (doi:10.1038/nature05342).
- Cisowski SM, Collinson DW, Runcorn SK, Stephenson A, and Fuller M (1983) A review of lunar paleointensity data and implications for the origin of lunar magnetism. *Journal of Geophysical Research Supplement* 88: A691–A704.
- Clarke JT, Ajello J, Ballester GE, et al. (2002) Ultraviolet emissions from the magnetic footprints of Io, Ganymede, and Europa on Jupiter. *Nature* 415: 997–1000.
- Clarke JT, Ballester G, Trauger J, et al. (1996) Far-ultraviolet imaging of Jupiter's aurora and the Io 'footprint'. *Science* 274: 404–409.
- Clarke JT, Grodent D, Cowley S, et al. (2005) Jupiter's aurorae. In: *Jupiter: The Planet, Satellites, and Magnetosphere*, pp. 639–671. Cambridge, UK: Cambridge University Press.
- Clarke JT, Prange R, and Ballester GE (1995) HST far-ultraviolet imaging of Jupiter during the impacts of Comet Shoemaker-Levy-9. *Science* 267(5202): 1302–1307.
- Colin L (1983) Basic facts about Venus. In: Hunten DM, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 10–26. Tucson, AZ: University of Arizona Press.
- Connerney JEP (1981) The magnetic field of Jupiter: A generalized inverse approach. *Journal of Geophysical Research* 86: 7679–7693.
- Connerney JEP (1992) Doing more with Jupiter's magnetic field. In: Rucker HO, Bauer SJ, and Kaiser ML (eds.) *Planetary Radio Emissions III*, pp. 13–33. Vienna: Austrian Academy of Science.
- Connerney JEP and Acuña MH (1982) Jovimagnetic secular variation. *Nature* 297: 313–315.
- Connerney JEP and Desch MD (1992) Comment on 'Evidence of Saturn's magnetic field anomaly from Saturnian kilometric radiation high frequency limit' by Galopéau et al. *Journal of Geophysical Research* 97: 8713–8717.
- Connerney JEP and Ness NF (1988) Mercury's magnetic field and interior. In: Matthews MS, Chapman L, and Vilas F (eds.) *Mercury*, pp. 494–513. Tucson, AZ: University of Arizona press.
- Connerney JEP, Acuña MH, and Ness NF (1981) Modeling the Jovian current sheet and inner magnetosphere. *Journal of Geophysical Research* 86: 8370–8384.
- Connerney JEP, Acuña MH, and Ness NF (1982a) Voyager 1 assessment of Jupiter's planetary magnetic field. *Journal of Geophysical Research* 87: 3623–3627.
- Connerney JEP, Ness NF, and Acuña MH (1982b) Zonal harmonic model of Saturn's magnetic field from Voyager 1 and 2 observations. *Nature* 298: 44–46.
- Connerney JEP, Acuña MH, and Ness NF (1984a) The Z_3 model of Saturn's magnetic field and the Pioneer 11 Vector Helium magnetometer observations. *Journal of Geophysical Research* 89: 7541–7544.
- Connerney JEP, Davis L, Jr., and Chenette DL (1984b) Magnetic field models. In: Gehrels T and Matthews MS (eds.) *Saturn*, pp. 354–377. Tucson, AZ: University of Arizona Press.
- Connerney JEP (1986) Magnetic connection for Saturn's rings and atmosphere. *Geophysical Research Letters* 13(8): 773–776.
- Connerney JEP, Acuña MH, and Ness NF (1987) The magnetic field of Uranus. *Journal of Geophysical Research* 92(A13): 15329–15336.
- Connerney JEP, Acuña MH, and Ness NF (1991) The magnetic field of Neptune. *Journal of Geophysical Research* 96: 19023–19042.
- Connerney JEP (1993) Magnetic fields of the outer planets. *Journal of Geophysical Research* 98(E10): 18659–18679.
- Connerney JEP, Baron R, Satoh T, and Owen T (1993) Images of excited H_{3+} at the Foot of the Io Flux Tube in Jupiter's atmosphere. *Science* 262: 1035–1038.
- Connerney JEP, Acuña MH, and Ness NF (1996) Octupole model of Jupiter's magnetic field from Ulysses observations. *Journal of Geophysical Research* 101: 27453–27458.
- Connerney JEP, Acuña MH, Wasilewski PJ, et al. (1999) Magnetic lineations in the ancient crust of Mars. *Science* 284: 794–798.
- Connerney JEP and Satoh T (2000) The H_{3+} ion: A remote diagnostic of the Jovian magnetosphere. *Philosophical Transactions of the Royal Society of London A* 358: 2471–2483.
- Connerney JEP, Acuña MH, Wasilewski PJ, et al. (2001) The global magnetic field of Mars and implications for crustal evolution. *Geophysical Research Letters* 28: 4015–4018.
- Connerney JEP, Acuña MH, Ness NF, and Satoh T (1998) New models of Jupiter's magnetic field constrained by the Io flux tube footprint. *Journal of Geophysical Research-Space Physics* 103(A6): 11929–11939.
- Connerney JEP, Acuña MH, Ness NF, Spohn T, and Schubert G (2004) Mars crustal magnetism. *Space Science Reviews* 111(1–2): 1–32.
- Connerney JEP, Acuña MH, Ness NF, et al. (2005) Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences* 102(42): 14970–14975.
- Cowley SWH, Wright DM, Bunce EJ, et al. (2006) Cassini observations of planetary-period magnetic field oscillations in Saturn's magnetosphere: Doppler shifts and phase motion. *Geophysical Research Letters* 33(7): L07104.
- Cowling TG (1933) The magnetic field of sunspots. *Monthly Notices of the Royal Astronomical Society* 94: 39–48.
- Cowling TG (1957) The dynamo maintenance of steady magnetic fields. Q. *Journal of Applied Mathematics* 10: 129–36.
- Crary FJ and Baggenal F (1998) Remanent ferromagnetism and the interior structure of Ganymede. *Journal of Geophysical Research* 103: 25757–25774.
- Crisp JA, Adler M, Matijevic JR, Squyres SW, Arvidson RE, and Kass DM (2003) Mars Exploration Rover mission. *Journal of Geophysical Research* 108(E12): 8061.
- Curtis SA and Ness NF (1986) Magnetospheric balance in planetary dynamos: Predictions for Neptune's magnetosphere. *Journal of Geophysical Research* 91: 11003–11008.
- Davis L, Jr. and Smith EJ (1976) The Jovian magnetosphere and magnetopause. In: McCormac BM (ed.) *Magnetospheric Particles and Fields*, pp. 301–310. Hingham, MA: Reidel D.
- Davis L, Jr. and Smith EJ (1985) Comments on 'The Z_3 model of Saturn's magnetic field and the Pioneer 11 Vector Helium

- magnetometer observations'. By: Connerney, Acuña, Ness. *Journal of Geophysical Research* 90: 4461–4464.
- Davis L, Jr. and Smith EJ (1986) New models of Saturn's magnetic field using Pioneer 11 Vector Helium magnetometer data. *Journal of Geophysical Research* 91: 1373–1380.
- de Pater I (1981) Radio maps of Jupiter's radiation belts and planetary disc at $\lambda = 6$ cm. *Astronomy and Astrophysics* 93: 370–381.
- Desch MD (1982) Evidence for solar-wind control of Saturn radio-emission. *Journal of Geophysical Research – Space Physics* 87(1): 4549–45.
- Desch MD, Connerney JEP, and Kaiser ML (1986) The rotation period of Uranus. *Nature* 322(6074): 42–43.
- Desch MD and Kaiser ML (1981) Voyager measurement of the rotation rate of Saturn's magnetic field. *Geophysical Research Letters* 8: 253–256.
- Desch MD and Rucker HO (1983) The relationship between Saturn kilometric radiation and the solar wind. *Journal of Geophysical Research* 88: 8999–9006.
- Desch MD, Kaiser ML, Zarka P, et al. (1991) Uranus as a radio source. In: Bergstrahl JT, Miner E, and Matthews MS (eds.) *Uranus*, pp. 894–925. Tucson, AZ: University of Arizona Press.
- Dessler AJ (1983) Coordinate systems. In: Dessler AJ (ed.) *Physics of the Jovian Magnetosphere*, pp. 498–504. New York: Cambridge University Press.
- Dougherty MK, Achilleos N, Andre N, et al. (2005) Cassini magnetometer observations during Saturn orbit insertion. *Science* 307(5713): 1266–1270.
- Dougherty MK, Balogh A, Southwood DJ, and Smith EJ (1996) Ulysses assessment of the Jovian planetary field. *Journal of Geophysical Research – Space Physics* 101(A11): 24929–24941.
- Dougherty MK, Kellock S, Southwood DJ, et al. (2004) The Cassini Magnetic Field Investigation. *Space Science Reviews* 114(1–4): 331–383.
- Elphic RC and Russell CT (1978) On the apparent source depth of planetary magnetic fields. *Geophysical Research Letters* 5: 211–214.
- Elsasser WM (1946) Induction effects in terrestrial magnetism. Part II: The secular variation. *Physical Review* 70: 202–212.
- Frawley JJ and Taylor PT (2004) Paleo-pole positions from martian magnetic anomaly data. *Icarus* 172(2): 316–327.
- Fuller M, Rose F, and Wasilewski PJ (1974) Preliminary results of an experimental study of the magnetic effects of shocking lunar soil. *The Moon* 9: 57–61.
- Galopéau P, Ortega-Molina A, and Zarka P (1991) Evidence of Saturn's magnetic field anomaly from Saturnian kilometric radiation high frequency limit. *Journal of Geophysical Research* 96: 14129–14140.
- Galopéau P, Zarka P, and LeQuéau D (1995) Source location of Saturn's kilometric radiation: The Kelvin-Helmholtz instability hypothesis. *Journal of Geophysical Research* 100: 26397–26410.
- Galopéau PHM and Lecacheux A (2000) Variations in Saturn's radio rotation period measured at kilometric wavelengths. *Journal of Geophysical Research* 105: 13089–13101.
- Giampieri G and Balogh A (2002) Mercury's thermoelectric dynamo model revisited. *Planetary Space Science* 50: 757–762.
- Giampieri G and Dougherty MK (2004) Rotation rate of Saturn's interior from magnetic field observations. *Geophysical Research Letters* 31(16): L16701.
- Giampieri G, Dougherty MK, Russell CT, and Smith EJ (2005) Reply to comment by Kaiser ML et al. on 'Rotation rate of Saturn's interior from magnetic field observations'. *Geophysical Research Letters* 32(2): L02202.
- Giampieri G, Dougherty MK, Smith EJ, and Russell CT (2006) A regular period for Saturn's magnetic field that may track its internal rotation. *Nature* 441(7089): 62–64.
- Gold RE, Solomon SC, McNutt RL, et al. (2001) The MESSENGER mission to Mercury: Scientific payload. *Planetary and Space Science* 49: 1467–1479.
- Goldreich P and Lynden-Bell D (1969) Io a Jovian unipolar inductor. *Astrophysical Journal* 156(1P1): 59–78.
- Goldreich P and Peale SJ (1966) Spin-orbit coupling in the solar system. *Astronomical Journal* 71: 425–438.
- Graf JE, Zurek RW, Eisen HJ, Jai B, Johnston MD, and DePaula R (2005) The Mars Reconnaissance Orbiter Mission. *Acta Astronautica* 57(2–8): 566–578.
- Gurnett DA, Kurth WS, Roux A, Bolton SJ, and Kennel CF (1996) Evidence for a magnetosphere at Ganymede from Plasma wave observations by the Galileo spacecraft. *Nature* 384: 535–537.
- Gurnett DA, Kurth WS, Hospodarsky GB, et al. (2005) Radio and plasma wave observations at Saturn from Cassini's approach and first orbit. *Science* 307: 1255–1259.
- Halekas JS, Mitchel DL, Lin RP, et al. (2001) Mapping of crustal magnetic anomalies on the lunar near side by the Lunar Prospector electron reflectometer. *Journal of Geophysical Research* 106(E11): 27841–27852.
- Halekas J, Mitchell DL, Lin RP, Hood LL, Acuña MH, and Binder AB (2002) Evidence for negative charging of the lunar surface in shadow. *Geophysical Research Letters* 29: 77–81.
- Hide R (1967) On the dynamics of Jupiter's interior and the origin of his magnetic field. In: Hindmarsh W, Lowes FJ, Roberts PH, and Runcorn SK (eds.) *Magnetism and the Cosmos*, pp. 378–395. Edinburgh: Oliver and Boyd.
- Hide R (1988) Towards the interpretation of Uranus's eccentric magnetic field. *Geophysical and Astrophysical Fluid Dynamics* 44: 207–209.
- Hood LL and Vickery A (1984) Magnetic field amplification and generation in hypervelocity meteoroid impacts with application to lunar paleomagnetism. *Journal of Geophysical Research* 89: C211–C223.
- Hood LL and Huang Z (1991) Formation of magnetic anomalies antipodal to lunar impact basins – Two dimensional model calculations. *Journal of Geophysical Research* 96: 9837–9846.
- Hood LL and Zakharian A (2001) Mapping and modeling of magnetic anomalies in the northern polar region of Mars. *Journal of Geophysical Research-Planets* 106(E7): 14601–14619.
- Hood LL, Zakharian A, Halekas JS, et al. (2001) Initial mapping and interpretation of Lunar crustal magnetic anomalies using Lunar Prospector magnetometer data. *Journal of Geophysical Research* 106(E11): 27825–27839.
- Hubbard WB and MacFarlane J (1980) Structure and evolution of Uranus and Neptune. *Journal of Geophysical Research* 85: 225–234.
- Hubbard WB and Marley MS (1989) Optimized Jupiter, Saturn, and Uranus interior models. *Icarus* 78: 102–118.
- Jackson DJ and Beard DB (1977) Magnetic-field of Mercury. *Journal of Geophysical Research-Space Physics* 82(19): 2828–2836.
- Johnson TV, Yeates CM, and Young R (1992) Space Science Reviews volume on Galileo Mission overview. *Space Science Reviews* 60: 3–21.
- Jurdy DM and Stefanick M (2004) Vertical extrapolation of Mars magnetic potentials. *Journal of Geophysical Research* 109(E10): E10005 (doi:10.1029/2004JE002277).
- Kaiser ML and Desch MD (1982) Saturn kilometric radiation: Source locations. *Journal of Geophysical Research* 87: 4555–4559.
- Kaiser ML, Desch MD, Kurth WS, et al. (1984) Saturn as a radio source. In: Gehrels T and Matthews MS (eds.) *Saturn*, pp. 378–415. Tucson, AZ: University of Arizona Press.
- Kaiser ML, Desch MD, Acuña MH, Ness NF, and Connerney JEP (2005) Comment on 'Rotation rate of

- Saturn's interior from magnetic field observations'. *Geophysical Research Letters* 32(2): L02201 (doi:10.1029/2004GL021398).
- Kargel JS (2006) Enceladus: Cosmic gymnast, volatile miniworld. *Science* 311: 1389–1391.
- Khurana KK, Kivelson MG, Stevenson DJ, et al. (1998) Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395: 777–780.
- Kirk RL and Stevenson DJ (1987) Hydromagnetic constraints on deep zonal flow in the giant planets. *Astrophysical Journal* 316: 836–846.
- Kivelson MG (2006) Does Enceladus govern magnetospheric dynamics at Saturn? *Science* 311: 1391–1392.
- Kivelson MG, Khurana KK, Russell C, et al. (1996) Discovery of Ganymede's magnetic field by the Galileo spacecraft. *Nature* 384: 537–541.
- Kivelson MG, Khurana KK, Coroniti FV, et al. (1997) The magnetic field and magnetosphere of Ganymede. *Geophysical Research Letters* 24: 2155–2158.
- Kivelson MG, Khurana KK, Stevenson DJ, et al. (1999) Europa and Callisto: Induced or intrinsic fields in a periodically varying plasma environment. *Journal of Geophysical Research* 104: 4609–4626.
- Kivelson MG, Khurana KK, Russell CT, Volwerk M, Walker RJ, and Zimmer C (2000) Galileo magnetometer measurements: A stronger case for a subsurface ocean at Europa. *Science* 289: 1340–1343.
- Kivelson MG, Khurana K, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.
- Kurth WS, Lecacheux A, Averkamp TF, Groene JB, and Gurnett DA (2007) A Saturnian longitude system based on a variable kilometric radiation period. *Geophysical Research Letters* 34(2): L02201.
- Langel RA (1987) The main field. In: Jacobs JA (ed.) *Geomagnetism*, vol. 1, pp. 249–513. New York: Academic.
- Langel RA and Estes RH (1982) A geomagnetic field spectrum. *Geophysical Research Letters* 9: 250–253.
- Langel RA and Estes RH (1985) The near-Earth magnetic field at 1980 determined from MAGSAT data. *Journal of Geophysical Research* 90: 2495–2509.
- Langel RA, Ousely G, Berbert J, Murphy J, and Settle M (1982) The MAGSAT mission. *Geophysical Research Letters* 9: 243–245.
- Langlais B, Purucker ME, and Mandea M (2004) Crustal magnetic field of Mars. *Journal of Geophysical Research - Planets* 109(E2): E02008.
- Lillis RJ, Mitchel DL, Lin RP, Connerney JEP, and Acuña MH (2004) Mapping crustal magnetic fields at Mars using electron reflectometry. *Geophysical Research Letters* 31: L15702.
- Lin RP, Mitchel DL, Curtis DW, et al. (1998) Lunar surface magnetic fields and their interaction with the solar wind: Results from Lunar Prospector. *Science* 281: 1480–1484.
- Lin RP, Anderson KA, and Hood LL (1988) Lunar surface magnetic field concentrations antipodal to young large impact basins. *Icarus* 74: 529–541.
- Lortz D (1972) A simple stationary dynamo model. *Zeitschrift für Naturforschung* 27: 1350–1354.
- Lowes FJ (1974) Spatial power spectrum of the main geomagnetic field and extrapolation to the core. *Geophysical Journal of the Royal Astronomical Society* 36: 717–730.
- Mayhew MA and Estes RH (1983) Equivalent source modeling of the core magnetic-field using MAGSAT data. *Journal of Geomagnetism and Geoelectricity* 35(4): 119–130.
- Mitchell DL, Lillis RJ, Lin RP, Connerney JEP, and Acuña MH (2007) A global map of Mars' crustal magnetic field based on electron reflectometry. *Journal of Geophysical Research - Planets* 112(E1): E01002.
- Mitchell AC and Nellis WJ (1982) Equation of state and electrical conductivity of water and ammonia shocked to the 100 GPa (1 Mbar) pressure range. *Journal of Chemical Physics* 76: 6273–6281.
- Mitchel DL, Halekas JS, Lin RP, Frey S, Hood LL, and Acuña MH, and Binder A (2007) Global mapping of Lunar crustal fields by Lunar Prospector. *Journal of Geophysical Research* (in press).
- Morris D and Berge GL (1962) Measurement of the polarization and angular extent of the decimeter radiation from Jupiter. *Astrophysical Journal* 136: 276–282.
- Nellis WJ, Hamilton DC, Holmes NC, et al. (1988) The nature of the interior of Uranus based on studies of planetary ices at high dynamic pressure. *Science* 240: 779–781.
- Ness NF (1979a) The magnetic fields of Mercury, Mars and Moon. *Annual Review of the Earth and Planetary Sciences* 7: 249–288.
- Ness NR (1979b) Magnetic-field of Mercury. *Physics of the Earth and Planetary Interiors* 20(2–4): 209–217.
- Ness NF, Behannon KW, Lepping RP, Whang YC, and Schatten KH (1974a) Magnetic field observations near Mercury: Preliminary results from Mariner 10. *Science* 183: 130–306.
- Ness NF, Behannon KW, Lepping RP, Whang YC, and Schatten KH (1974b) Magnetic field observations near Venus: Preliminary results from Mariner 10. *Science* 185: 151–160.
- Ness NF, Behannon KW, Lepping RP, and Whang YC (1975) Magnetic field of Mercury confirmed. *Nature* 255: 204–205.
- Ness NF, Behannon KW, Lepping RP, and Whang YC (1976) Observations of Mercury's magnetic field. *Icarus* 28: 479–488.
- Ness NF, Acuña MH, Lepping RP, Burlaga LF, Behannon KW, and Neubauer FM (1979a) Magnetic field studies at Jupiter by Voyager 1: Preliminary results. *Science* 204: 982–987.
- Ness NF, Acuña MH, Lepping RP, Burlaga LF, Behannon KW, and Neubauer FM (1979b) Magnetic field studies at Jupiter by Voyager 2: Preliminary results. *Science* 206: 966–972.
- Ness NF, Acuña MH, Behannon KW, et al. (1981) Magnetic field studies by Voyager 1: Preliminary results at Saturn. *Science* 212: 211–217.
- Ness NF, Acuña MH, Behannon KW, et al. (1982) Magnetic field studies by Voyager 2: Preliminary results at Saturn. *Science* 215: 558–563.
- Ness NF, Acuña MH, Behannon KW, et al. (1986) Magnetic fields at Uranus. *Science* 233: 85–89.
- Ness NF, Acuña MH, Burgala LF, Connerney JEP, Lepping RP, and Neubauer FM (1989) Magnetic fields at Neptune. *Science* 246: 1473–1478.
- Neubauer FM (1998) The sub-Alfvénic interaction of the Galilean satellites with the Jovian magnetosphere. *Journal of Geophysical Research* 103(E9): 19843–19866.
- Ng KH and Beard DB (1979) Possible displacement of Mercury dipole. *Journal of Geophysical Research-Space Physics* 84(NA5): 2115–2117.
- Nimmo F (2000) Dike intrusion as a possible cause of linear Martian magnetic anomalies. *Geology* 28: 391–394.
- Nimmo F and Stevenson D (2000) Influence of early plate tectonics on the thermal evolution and magnetic field of Mars. *Journal of Geophysical Research* 105: 11969–11976.
- Northrop TG and Connerney JEP (1987) A micrometeorite erosion model and the age of Saturn's rings. *Icarus* 70: 124–137.
- Parker EN (1955) Hydromagnetic dynamo models. *Astrophysical Journal* 122: 293–314.
- Parker EN (1969) The occasional reversal of the geomagnetic field. *Astrophysical Journal* 158: 815–827.

- Parker EN (1971) The generation of magnetic fields in astrophysical bodies. Part IV: The solar and terrestrial dynamos. *Astrophysical Journal* 164: 491–509.
- Parker RL (2003) Ideal bodies for Mars magnetics. *Journal of Geophysical Research* 108(E1): 5006 (doi:10.1029/2001JE00760).
- Pettengill GH and Dyce RB (1965) A radar determination of the rotation of the planet Mercury. *Nature* 206: 1240–1241.
- Phillips JL and Russell CT (1987) Upper limit on the intrinsic magnetic field of Venus. *Journal of Geophysical Research* 92: 2253–2263.
- Podolak M and Reynolds RT (1981) On the structure and composition of Uranus and Neptune. *Icarus* 46: 40–50.
- Podolak M and Reynolds RT (1987) The rotation rate of Uranus, its internal structure, and the process of planetary accretion. *Icarus* 70: 31–36.
- Podolak M, Reynolds RT, and Young R (1990) Post Voyager comparisons of the interiors of Uranus and Neptune. *Geophysical Research Letters* 17: 1737–1740.
- Podolak M, Hubbard WB, and Stevenson DJ (1991) Models of Uranus' interior and magnetic field. In: Bergstrahl JT, Miner ED, and Matthews MS (eds.) *Uranus*, pp. 29–61. Tucson, AZ: University of Arizona Press.
- Porco CC and Danielson GE (1982) The periodic variation of spokes in Saturn's rings. *Astronomical Journal* 87: 826–833.
- Prange R, Rego D, Southwood D, Zarka P, Miller S, and Ip W-H (1996) Rapid energy dissipation and variability of the Io–Jupiter electrodynamic circuit. *Nature* 379: 232–235.
- Purucker M, Ravat D, Frey H, Sabaka T, and Acuña MH (2000) An altitude-normalized magnetic map of Mars and its interpretation. *Geophysical Research Letters* 27(16): 2449–2452.
- Radhakrishnan V and Roberts JA (1960) Polarization and angular extent of the 960 Mc/sec radiation from Jupiter. *Physical Review Letters* 4: 493–494.
- Rädler KH and Ness NF (1990) The symmetry properties of planetary magnetic fields. *Journal of Geophysical Research* 95: 2311–2318.
- Riddle AC and Warwick JW (1976) Redefinition of System III longitude. *Icarus* 27: 457–459.
- Riedler W, Mohlmann D, Oraevsky VN, et al. (1989) Magnetic fields near Mars – 1st results. *Nature* 341(6243): 604–607.
- Runcorn SK (1975) An ancient lunar magnetic dipole field. *Nature* 253: 701–703.
- Runcorn SK (1975) On the interpretation of lunar magnetism. *Physics of the Earth and Planetary Interiors* 10: 327–335.
- Russell CT (1978a) The magnetic field of Mars: Mars 3 evidence re-examined. *Geophysical Research Letters* 5: 81–84.
- Russell CT (1978b) The magnetic field of Mars: Mars 5 evidence re-examined. *Geophysical Research Letters* 5: 85–88.
- Russell CT (1979) The interaction of the solar wind with Mars, Venus and Mercury. In: Kennel CF (ed.) *Solar System Plasma Physics*, pp. 208–252. Amsterdam: Elsevier.
- Russell CT, Elphic RC, Luhmann JG, and Slavin JA (1980) On the search for an intrinsic magnetic field at Venus. *Proceedings of the Lunar and Planetary Science Conference* 11: 1897–1906.
- Russell CT and Vaisberg O (1983) The interaction of the solar wind with Venus. In: Hunten M, Colin L, Donahue TM, and Moroz VI (eds.) *Venus*, pp. 873–940. Tucson, AZ: University of Arizona Press.
- Ruzmaikin AA and Stachenko SV (1991) On the origin of Uranus and Neptune magnetic fields. *Icarus* 93: 82–87.
- Sabaka TJ, Olsen N, and Langel RA (2002) A comprehensive model of the quiet-time, near-Earth magnetic field: phase 3. *Geophysical Journal International* 151: 32–68.
- Saunders RS, Arvidson RE, Badhwar GD, et al. (2004) 2001 Mars Odyssey mission summary. *Space Science Reviews* 110(1–2): 1–36.
- Sandel BR, Shemansky DE, Broadfoot AL, et al. (1982) Extreme ultraviolet observations from the Voyager 2 encounter with Saturn. *Science* 215: 548–553.
- Schubert G and Spohn T (1990) Thermal history of Mars and the sulfur content of its core. *Journal of Geophysical Research* 95: 14095–14104.
- Schubert G, Solomon SC, Turcotte DL, Drake MJ, and Sleep N (1992) Origin and thermal evolution of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 147–183. Tucson, AZ: University of Arizona Press.
- Schubert G, Russell CT, and Moore WB (2000) Geophysics – Timing of the Martian dynamo. *Nature* 408: 666–667.
- Schubert G, Chan KH, Xinhao L, and Zhang K (2004) Planetary dynamos: Effects of electrical conducting flows overlying turbulent regions of magnetic field generation. *Icarus* 172: 305–315.
- Schlitz PH and Lutz AB (1988) Polar wandering on Mars. *Icarus* 73(1): 91–141.
- Schulz M and Paulikas GA (1990) Planetary magnetic fields: A comparative view. *Advances in Space Research* 10(1): 55–64.
- Simpson JA, Bastian TS, Chenette DL, McKibben RB, and Pyle KR (1980) The trapped radiations of Saturn and their absorption by satellites and rings. *Journal of Geophysical Research* 85: 5731–5762.
- Sleep NH (1990) Hotspots and mantle plumes – Some phenomenology. *Journal of Geophysical Research-Solid Earth and Planets* 95(B5): 6715–6736.
- Sleep NH (1994) Martian plate tectonics. *Journal of Geophysical Research* 99: 5639–5655.
- Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1503.
- Smith DE, Zuber MT, and Neumann GA (2001) Seasonal variations of snow depth on Mars. *Science* 294(5549): 2141–2146.
- Smith EJ, Davis L, Jr., Coleman PJ, and Sonnet CP (1963) Mariner II: Preliminary reports on measurements of Venus magnetic field. *Science* 139: 909–910.
- Smith EJ, Davis L, Jr., Coleman PJ, and Sonnet CP (1965) Magnetic measurements near Venus. *Journal of Geophysical Research* 70: 1571–1586.
- Smith EJ, Davis L, Jr., Coleman PJ, and Jones DE (1965) Magnetic field measurements near Mars. *Science* 149: 1241–1242.
- Smith EJ, Davis L, Jr., Jones DE, et al. (1974) Magnetic field of Jupiter and its interaction with the solar wind. *Science* 183: 305–306.
- Smith EJ, Davis L, Jr., Jones DE, et al. (1975) Jupiter's magnetic field, magnetosphere, and its interaction with the solar wind: Pioneer 11. *Science* 188: 451–455.
- Smith EJ, Connor BV, and Foster GT, Jr. (1975) Measuring the magnetic fields of Jupiter and the outer solar system. *IEEE Transactions on Magnetics*, MAG-11 962–980.
- Smith EJ, Davis L, Jr., and Jones DE (1976) Jupiter's magnetic field and magnetosphere. In: Gehrels T (ed.) *Jupiter*, pp. 788–829. Tucson, AZ: University of Arizona Press.
- Smith EJ, Davis L, Jr., Jones DE, et al. (1980a) Saturn's magnetic field and magnetosphere. *Science* 207: 407–410.
- Smith EJ, Davis L, Jr., Jones DE, et al. (1980b) Saturn's magnetosphere and its interaction with the solar wind. *Journal of Geophysical Research* 85: 5655–5674.
- Smoluchowski R (1975) Jupiter's molecular hydrogen layer and the magnetic field. *Astrophysical Journal Letters* 200: L119–L121.
- Smoluchowski R and Torbett M (1981) Can magnetic fields be generated in the icy mantles of Uranus and Neptune? *Icarus* 48: 146–148.

- Solomon SC (1976) Some aspects of core formation in Mercury. *Icarus* 28: 509–521.
- Solomon SC, McNutt RL, Jr., Gold RE, et al. (2001) The Messenger mission to Mercury: Scientific objectives and implementation. *Planetary and Space Science* 49: 1445–1465.
- Spohn T (1991) Mantle differentiation and thermal evolution of Mars, Mercury, and Venus. *Icarus* 90(2): 222–236.
- Spohn T, Acuña MH, Breuer D, et al. (2001) Geophysical constraints on the evolution of Mars. *Space Science Reviews* 96: 231–262.
- Spohn T, Sohl F, and Breuer D (1998) Mars. *Astronomy and Astrophysics Review* 8: 181–235.
- Sprenke KF (2005) Martian magnetic paleopoles: A geostatistical approach. *Geophysical Research Letters* 32(9): L09201.
- Sprenke KF and Baker LL (2000) Magnetization, paleomagnetic poles, and polar wander on Mars. *Icarus* 147: 26–34.
- Stegman DR, Jellinek AM, Zatman SA, Baumgardner JR, and Richards MA (2003) An early lunar core dynamo driven by thermochemical mantle convection. *Nature* 421: 143–146.
- Stevenson DJ (1980) Saturn's luminosity and magnetism. *Science* 208: 746–748.
- Stevenson DJ (1982) Reducing the non-axisymmetry of a planetary dynamo and an application to Saturn. *Geophysical and Astrophysical Fluid Dynamics* 21: 113–127.
- Stevenson DJ (1983) Planetary magnetic fields. *Reports on Progress in Physics* 46: 555–620.
- Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11.
- Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.
- Torbett M and Smoluchowski R (1979) The structure and magnetic field of Uranus. *Geophysical Research Letters* 6: 675–676.
- Van Allen JA, Thomsen MF, Randall BA, Raider RL, and Grosskreutz CL (1980) Saturn's magnetosphere, rings, and inner satellites. *Science* 207: 415–421.
- Warwick JW, Evans DR, Peltzer GR, et al. (1989) Voyager planetary radio astronomy at Neptune. *Science* 246(4936): 1498–1501.
- Whang YC (1977) Magnetospheric magnetic-field of Mercury. *Journal of Geophysical Research-Space Physics* 82(7): 1024–1030.
- Warwick JW (1963) The position and sign of Jupiter's magnetic moment. *Astrophysical Journal* 137: 1317–1318.
- Warwick JW, Evans DR, Peltzer GR, et al. (1989) Voyager planetary radio astronomy at Neptune. *Science* 246(4936): 1498–1501.
- Whang YC (1977) Magnetospheric magnetic-field of Mercury. *Journal of Geophysical Research-Space Physics* 82(7): 1024–1030.
- Zimmer C, Khurana KK, and Kivelson MG (2000) Subsurface oceans on Europa and Callisto: Constraints from Galileo magnetometer observations. *Icarus* 147: 329–347.

Relevant Website

<http://mgs-mager.gsfc.nasa.gov> – Mars Global Surveyor.

10.08 Planetary Dynamos

F. H. Busse, Physikalisches Institut der Universität Bayreuth, Bayreuth, Germany

R. Simitev, University of Glasgow, Glasgow, UK

© 2007 Elsevier B.V. All rights reserved.

10.08.1	Historical Introduction	281
10.08.2	General Remarks on the Dynamo Theory of Planetary Magnetism	282
10.08.3	Mathematical Formulation of the Problem of Spherical Dynamos	283
10.08.4	Convection in Rotating Spherical Shells	285
10.08.5	Convection-Driven Dynamos	288
10.08.6	Applications to Planetary Dynamos	292
10.08.6.1	General Considerations	292
10.08.6.2	Mercury	294
10.08.6.3	Venus	295
10.08.6.4	Mars and Moon	295
10.08.6.5	Jupiter	295
10.08.6.6	Saturn	295
10.08.6.7	Uranus and Neptune	295
10.08.6.8	Ganymede and Other Satellites	296
10.08.7	Concluding Remarks	296
	References	297

10.08.1 Historical Introduction

While the Earth's magnetism has been studied for centuries starting with the first scientific monograph of [Gilbert \(1600\)](#), the question of the magnetism of other planets had received scant attention until recently because of the lack of relevant observations. Only in 1955 clear evidence for the existence of a planetary magnetic field other than the geomagnetic one was obtained through the observation of the Jovian decametric radio waves ([Burke and Franklin, 1955](#)). Since it had been more or less accepted until the end of the nineteenth century that geomagnetism arises from the remnant magnetization of the Earth similar properties may have been assumed for the other terrestrial planets and the Moon (cf. Chapter 8.03). This view lost its appeal, however, when it became evident that the Curie temperature is exceeded in the Earth below a depth of about 30 km. Ferromagnetic materials in the Earth's crust could thus explain only magnetic fields with a relatively short wavelength.

The current period of intense research on the magnetism of planets other than that of the Earth started with the first detailed measurement of Jupiter's magnetic field by the Pioneer 10 space

probe in 1973 and the discovery of Mercury's magnetism by Mariner 10 in 1974. In the early 1970s also the development of the theory of magnetohydrodynamic (MHD) dynamos had started in which the reaction of the Lorentz force of the generated magnetic field is taken into account in physically realistic configurations ([Childress and Soward, 1972; Busse, 1973; Soward, 1974](#)). Until that time dynamo theoreticians had focused their attention on the kinematic problem in which the possibility of growing magnetic fields driven by somewhat arbitrarily chosen velocity fields is considered. It must be remembered that only a few years earlier it had been demonstrated by [Backus \(1958\)](#) and [Herzenberg \(1958\)](#) in a mathematically convincing way that the homogeneous dynamo process of the generation of magnetic fields in a singly connected domain of an electrically conducting fluid is indeed possible. Doubts about the feasibility of this process which had first been proposed by [Larmor \(1919\)](#) had persisted after [Cowling \(1934\)](#) had proved that purely axisymmetric fields could not be generated in this way.

The complexity of the MHD dynamo problem described by the nonlinearly coupled Navier-Stokes equations and the equation of magnetic

induction had prevented progress in understanding planetary dynamos through analytical solutions. Only the advent of powerful enough computers in the 1990s has allowed to solve numerically the coupled three-dimensional (3-D) partial differential equations through forward integration in time. Even today and for the foreseeable future the limits of computer capacity will permit the exploration of only a small fraction of the parameter space of interest for the understanding of planetary dynamos.

In view of the difficulties of a rigorous theory of planetary dynamos, many attempts have been made to obtain simple similarity relationships which would fit the observed planetary magnetic moments as function of certain properties of the planets. Some early proponents have gone as far as claiming the existence of a ‘magnetic Bode’s law’ corresponding to a relationship between the magnetic moment and size or angular momentum of a planet in analogy to the Titius–Bode law for the radii of the orbits of the planets. Just as in the latter case, however, attempts to derive a magnetic Bode’s law from basic physical principles have failed.

Other proposals have taken into account physical forces. Since a common ingredient of planetary dynamos is the existence of a fluid part of the core with a sufficiently high electrical conductivity the latter parameter together with the core radius and the angular velocity of the planetary rotation usually enter the similarity relationships such as those proposed by Hide (1974), Busse (1976), Jacobs (1979), and Dolginov (1977). Malkus (1968, 1994) has argued for the precession of the Earth as the cause of geomagnetism and he and Vanyo (1984) have demonstrated through laboratory experiments that precession and tides may cause turbulent motions in fluid planetary cores. Laboratory experiments on the geodynamo are reviewed in Chapter 8.11. Dolginov (1977) proposed a scaling law for the precessional origin of all planetary magnetic fields. While a dynamo driven by turbulent flows caused by precession and tides cannot be easily excluded in the case of the Earth (Tilgner, 2005), it is much less likely in the case of other planets such as Uranus for which precessional torques are rather minute. Just as a common precessional origin of planetary magnetism is not regarded as feasible, so have all other proposed similarity relationships lost in appeal and are no longer seriously considered. We shall return, however, to scaling relationships based more directly on the basic equations in Section 10.08.6.

10.08.2 General Remarks on the Dynamo Theory of Planetary Magnetism

Since the proposal of the geodynamo as the cause of the Earth’s magnetism had been in doubt for a long time before 1958, numerous alternative proposals had been made in the literature. Among these only the possibility that thermoelectric currents may generate a planetary magnetic field is still discussed in the case of Mercury (Stevenson, 1987; Giampieri and Balogh, 2002). For a discussion of the failings of the various proposals for nondynamo origins of planetary magnetic fields, we refer to the papers of Bullard (1949) and Stevenson (1983). Although the dynamo hypothesis of the origin of planetary magnetism is not without difficulties, it is the only one considered seriously at the present time with the possible exception of the just mentioned case of Mercury.

Dynamos generally convert mechanical energy into magnetic one. In contrast to the technical dynamo which is characterized by a multiply connected region of high electrical conductivity, that is, it depends on an appropriate wiring, planetary dynamos are referred to as homogeneous dynamos since they operate in a singly connected domain of high electrical conductivity. Since flows in planetary cores with active dynamos are usually turbulent, the small-scale structure of the magnetic field is correspondingly chaotic. The large-scale structure, however, can be quite regular. One distinguishes ‘steady’ and oscillatory dynamos. The most famous example of the latter kind is the solar dynamo which exhibits a well-defined period of about 22 years. The geodynamo, on the other hand, is a ‘steady’ dynamo, even though it varies in its amplitude by a factor of 2 or more on the magnetic diffusion timescale and reverses its polarity on a longer timescale. A measure of the magnetic diffusion time is given by the decay time, $t_d = \sigma \mu r_0^2 / \pi^2$, of the magnetic field in the absence of fluid motions. Here σ and μ refer to the electrical conductivity and the magnetic permeability of the planetary core of radius r_0 . In the case of the Earth, the decay time is of the order of 20 000 years, but it may vary between a few hundred and a million years for other examples of planetary dynamos.

The theory of homogeneous dynamos is based on Maxwell’s equations for the magnetic flux density \mathbf{B} , the electric current density \mathbf{j} , and the electric field \mathbf{E} in the MHD approximation in which the displacement current is neglected. This approximation is highly

accurate as long as the fluid velocity is small compared to the velocity of light which is certainly the case for all planetary applications. These equations together with Ohm's law for a moving conductor are given by

$$\nabla \cdot \mathbf{B} = 0, \quad \frac{\partial}{\partial t} \mathbf{B} = -\nabla \times \mathbf{E} \quad [1a]$$

$$\nabla \times \mathbf{B} = \mu \mathbf{j}, \quad \mathbf{j} = \sigma(\mathbf{v} \times \mathbf{B} + \mathbf{E}) \quad [1b]$$

where μ is the magnetic permeability of the fluid and σ is its electrical conductivity. These 'pre-Maxwell' equations have the property that they are invariant with respect to a Galilei transformation, that is, the equations remain unchanged in a new frame of reference moving with the constant velocity vector \mathbf{V} relative to the original frame of reference. Indicating the variables of the new frame by a prime, we find

$$\mathbf{v}' = \mathbf{v} - \mathbf{V}, \quad \frac{\partial}{\partial t'} = \frac{\partial}{\partial t} + \mathbf{V} \cdot \nabla \quad [2a]$$

$$\mathbf{B}' = \mathbf{B}, \quad \mathbf{E}' = \mathbf{E} + \mathbf{V} \times \mathbf{B}, \quad \mathbf{j}' = \mathbf{j} \quad [2b]$$

This invariance is the basis for the combination in MHD of eqns [1] with the equations of hydrodynamics in their usual nonrelativistic form. It is remarkable that this invariance does not only hold with respect to a Galilei transformation, but with respect to a transformation to a rotating frame of reference as well. In that case \mathbf{V} is replaced by $\boldsymbol{\Omega} \times \mathbf{r}$ in eqns [2], but when $\partial/\partial t'$ is operating on any vector \mathbf{a} , the term $-\boldsymbol{\Omega} \times \mathbf{a}$ must be added on the right-hand side, since even a constant vector field becomes time dependent when seen from a rotating frame unless it is parallel to $\boldsymbol{\Omega}$.

Elimination of \mathbf{E} and \mathbf{j} from eqn [1] yields the equation of magnetic induction

$$\nabla \times \left(\frac{1}{\sigma\mu} \nabla \times \mathbf{B} \right) = \frac{\partial}{\partial t} \mathbf{B} + \nabla \times (\mathbf{v} \times \mathbf{B}) \quad [3]$$

which for a solenoidal velocity field \mathbf{v} and a constant magnetic diffusivity $\lambda \equiv (\sigma\mu)^{-1}$ can be further simplified:

$$\left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{B} - \lambda \nabla^2 \mathbf{B} = \mathbf{B} \cdot \nabla \mathbf{v} \quad [4]$$

This equation has the form of a heat equation with the magnetic field line stretching term on the right-hand side acting as a heat source. This interpretation is especially useful for the dynamo problem. In order that a magnetic field \mathbf{B} may grow, the term on the right-hand side of [4] must overcome the effect of the magnetic diffusion term on the left-hand side. Using a typical velocity U and a typical length scale d , the

ratio of the two terms can be estimated by the magnetic Reynolds number Rm :

$$Ud/\lambda \equiv Rm \quad [5]$$

Only when Rm is of the order 1 or larger may growing magnetic fields become possible.

In the following we shall first consider the mathematical formulation of the problem of convection-driven dynamos in rotating spherical shells in a simple form in which only the physically most relevant parameters are taken into account. Before discussing dynamo solutions in Section 10.08.5 we shall briefly outline in Section 10.08.4 properties of convection in the absence of a magnetic field. Convection in the Earth core is reviewed in Chapter 8.05 and the theory of geodynamo in Chapter 8.03. Numerical solutions of the geodynamo problem are reviewed in Chapter 8.08. Applications to various planets and moons will be considered in Section 10.08.6 of this chapter and some concluding remarks are given in Section 10.08.7.

10.08.3 Mathematical Formulation of the Problem of Spherical Dynamos

A sketch of the geometrical configuration that will be considered is shown in Figure 1. For the equations

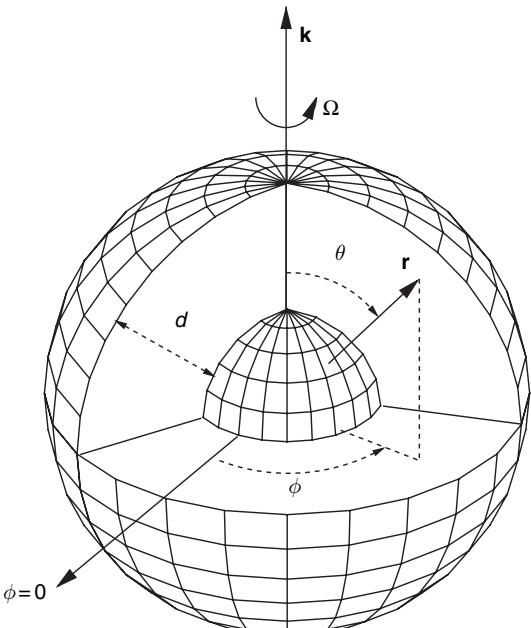


Figure 1 Geometrical configuration of the problem. A part of the outer spherical surface is removed to expose the interior of the shell to which the conducting fluid is confined.

describing convection-driven dynamos in the frame rotating with the angular velocity Ω we use a standard formulation which has also been used for a dynamo bench-mark (Christensen *et al.*, 2001). But we assume different scales and assume that a more general static state exists with the temperature distribution $T_S = T_0 - \beta d^2 r^2 / 2 + \Delta T \eta r^{-1} (1-\eta)^{-2}$, where η denotes the ratio of inner-to-outer radius of the spherical shell and d is its thickness. ΔT is the temperature difference between the boundaries in the special case $\beta=0$. In the case $\Delta T=0$, the static temperature distribution T_S corresponds to that of a homogeneously heated sphere with the heat source density proportional to the parameter β . The gravity field is given by $\mathbf{g} = -\gamma \mathbf{dr}$ where \mathbf{r} is the position vector with respect to the center of the sphere and r is its length measured in units of d .

In addition to d , the time d^2/ν , the temperature $\nu^2/\gamma\alpha d^4$, and the magnetic flux density $\nu(\mu\varrho)^{1/2}/d$ are used as scales for the dimensionless description of the problem where ν denotes the kinematic viscosity of the fluid, κ its thermal diffusivity, and ϱ its density. The Boussinesq approximation is used in that ϱ is assumed to be constant except in the gravity term where its temperature dependence given by $\alpha \equiv -(d\varrho/dT)/\varrho = \text{const.}$ is taken into account. The dimensionless equations of motion, the heat equation for the deviation Θ of the temperature field from the static distribution, and the equation of magnetic induction thus assume the form

$$\nabla^2 \mathbf{v} + \mathbf{B} \cdot \nabla \mathbf{B} + \mathbf{r}\Theta - \nabla\pi = P^{-1}(\partial_t \mathbf{v} + \mathbf{v} \cdot \nabla \mathbf{v}) + \tau \mathbf{k} \times \mathbf{v} \quad [6a]$$

$$\nabla \cdot \mathbf{v} = 0 \quad [6b]$$

$$\nabla^2 \Theta + [R_i + R_e \eta r^{-3} (1-\eta)^{-2}] \mathbf{r} \cdot \mathbf{v} = P(\partial_t + \mathbf{v} \cdot \nabla) \Theta \quad [6c]$$

$$P_m \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{B} - \nabla^2 \mathbf{B} = P_m \mathbf{B} \cdot \nabla \mathbf{v} \quad [6d]$$

and where \mathbf{k} is the unit vector in the direction of the axis of rotation where $\nabla\pi$ includes all terms that can be written as gradients. The Rayleigh numbers R_i and R_e , the Coriolis parameter τ , the Prandtl number P , and the magnetic Prandtl number P_m are defined by

$$R_i = \frac{\alpha\gamma\beta d^6}{\nu\kappa}, \quad R_e = \frac{\alpha\gamma\nabla T d^4}{\nu\kappa} \quad [7a]$$

$$\tau = \frac{2\Omega d^2}{\nu}, \quad P = \frac{\nu}{\kappa}, \quad P_m = \frac{\nu}{\lambda} \quad [7b]$$

For simplicity $R_e=0$ will be assumed unless indicated otherwise. The notation $R \equiv R_i$ will thus be used

in the following. The Prandtl number P has been added as an important parameter of the problem even though $P=1$ is often assumed with the argument that all effective diffusivities are equal in turbulent media. The effective diffusivities of scalar and vector quantities in turbulent fluid flow differ in general, however, and large differences in the corresponding molecular diffusivities will not be erased entirely in the turbulent case.

Since the velocity field \mathbf{v} as well as the magnetic flux density \mathbf{B} are solenoidal vector fields, the general representation in terms of poloidal and toroidal components can be used:

$$\mathbf{v} = \nabla \times (\nabla \Phi \times \mathbf{r}) + \nabla \Psi \times \mathbf{r} \quad [8a]$$

$$\mathbf{B} = \nabla \times (\nabla b \times \mathbf{r}) + \nabla g \times \mathbf{r} \quad [8b]$$

By multiplying the $(\text{curl})^2$ and the curl of eqn [6a] by \mathbf{r} , we obtain two equations for Φ and Ψ :

$$[(\nabla^2 - \partial_t) L_2 + \tau \partial_\phi] \nabla^2 \Phi + \tau \mathcal{Q} \Psi - L_2 \Theta = -\mathbf{r} \cdot \nabla \times [\nabla \times (\mathbf{v} \cdot \nabla \mathbf{v} - \mathbf{B} \cdot \nabla \mathbf{B})] \quad [9a]$$

$$[(\nabla^2 - \partial_t) L_2 + \tau \partial_\phi] \Psi - \tau \mathcal{Q} \Phi = \mathbf{r} \cdot \nabla \times (\mathbf{v} \cdot \nabla \mathbf{v} - \mathbf{B} \cdot \nabla \mathbf{B}) \quad [9b]$$

where ∂_t and ∂_ϕ denote the partial derivatives with respect to time t and with respect to the angle ϕ of a spherical system of coordinates r, θ, ϕ and where the operators L_2 and \mathcal{Q} are defined by

$$L_2 \equiv -r^2 \nabla^2 + \partial_r(r^2 \partial_r)$$

$$\mathcal{Q} \equiv r \cos \theta \nabla^2 - (L_2 + r \partial_r)(\cos \theta \partial_r - r^{-1} \sin \theta \partial_\theta)$$

The equations for b and g are obtained through the multiplication of eqn [6d] and of its curl by \mathbf{r} :

$$\nabla^2 L_2 b = P_m [\partial_t L_2 b - \mathbf{r} \cdot \nabla \times (\mathbf{v} \times \mathbf{B})] \quad [10a]$$

$$\nabla^2 L_2 g = P_m [\partial_t L_2 g - \mathbf{r} \cdot \nabla \times (\nabla \times (\mathbf{v} \times \mathbf{B}))] \quad [10b]$$

Either rigid boundaries with fixed temperatures as in the benchmark case (Christensen *et al.*, 2001),

$$\Phi = \partial_r(r\Phi) = \Psi = \Theta = 0 \quad \text{at} \quad r = r_i \equiv \eta/(1-\eta) \quad \text{and} \quad r = r_o = (1-\eta)^{-1} \quad [11]$$

or stress-free boundaries with fixed temperatures,

$$\Phi = \partial_{rr}^2 \Phi = \partial_r(\Psi/r) = \Theta = 0 \quad \text{at} \quad r = r_i \equiv \eta/(1-\eta) \quad \text{and} \quad r = r_o = (1-\eta)^{-1} \quad [12]$$

are often used. The latter boundary conditions are assumed in the following since they allow to cover numerically a larger region of the parameter space. The case $\eta = 0.4$ will be considered unless indicated otherwise. It provides a good compromise for the study of both, the regions inside and outside the tangent cylinder. The latter is the cylindrical surface touching the inner spherical boundary at its equator. For the magnetic field electrically insulating boundaries are used such that the poloidal function b must be matched to the function $b^{(e)}$ which describes the potential fields outside the fluid shell

$$g = b - b^{(e)} = \partial_r(b - b^{(e)}) = 0 \quad \text{at } r = r_i \text{ and } r = r_o \quad [13]$$

But computations for the case of an inner boundary with no-slip conditions and an electrical conductivity equal to that of the fluid have also been done. The numerical integration of eqns [2] together with boundary conditions [4] proceeds with the pseudo-spectral method as described by [Tilgner and Busse \(1997\)](#) and [Tilgner \(1999\)](#) which is based on an expansion of all dependent variables in spherical harmonics for the θ, ϕ -dependences, that is,

$$\Phi = \sum_{l,m} V_l^m(r, t) P_l^m(\cos \theta) \exp \{im\phi\} \quad [14]$$

and analogous expressions for the other variables, Ψ, Θ, b , and g . P_l^m denotes the associated Legendre functions. For the r -dependence expansions in Chebyshev polynomials are used.

For the computations to be reported in the following a minimum of 33 collocation points in the radial direction and spherical harmonics up to the order 64 have been used. But in many cases the resolution was increased to 49 collocation points and spherical harmonics up to the order 96 or 128.

It should be emphasized that the static state $\mathbf{v} = \mathbf{B} = \Theta = 0$ represents a solution of eqns [6] for all values of the Rayleigh numbers R_i and R_e , but this solution is unstable except for low or negative values of the latter parameters. Similarly, there exist solutions with $\mathbf{B} = 0$, but $\mathbf{v} \neq 0, \Theta \neq 0$, for sufficiently large values of either R_i or R_e or both, but, again, these solutions are unstable for sufficiently large values of P_m with respect to disturbances with $\mathbf{B} \neq 0$. Dynamo solutions as all solutions with $\mathbf{B} \neq 0$ are called are thus removed by at least two bifurcations from the basic static solution of the problem.

Finally, we present in [Table 1](#) a list of the most important parameters used in the following sections.

Table 1 Important dimensionless parameters

E	Average density of kinetic energy	[18]
E_P, E_t	Energy densities of poloidal and toroidal components of motion	[18]
η	Radius ratio of fluid shell	
Λ	Elsasser number	[20]
M	Average magnetic energy density	[19]
M_p, M_t	Energy densities of poloidal and toroidal components of the magnetic field	[19]
Nu	Nusselt number	[16]
P	Prandtl number	[7b]
P_m	Magnetic Prandtl number	[7b]
R	Rayleigh number	[7a]
R_c	Critical value of R for onset of convection	
R_m	Magnetic Reynolds number $= P_m \sqrt{2E}$	
τ	Coriolis parameter	[7b]

10.08.4 Convection in Rotating Spherical Shells

For an introduction to the problem of convection in spherical shells, we refer to the review of [Busse \(2002a\)](#). Convection tends to set in first outside the tangent cylinder in the form of thermal Rossby waves for which the Coriolis force is balanced almost entirely by the pressure gradient. The model of the rotating cylindrical annulus has been especially useful for the understanding of this type of convection. A rough idea of the dependence of the critical Rayleigh number R_c for the onset of convection on the parameters of the problem can be gained from the expressions derived from the annulus model ([Busse, 1970](#)):

$$R_c = 3 \left(\frac{P\tau}{1+P} \right)^{4/3} (\tan \theta_m)^{8/3} r_m^{-1/3} 2^{-2/3} \quad [15a]$$

$$m_c = \left(\frac{P\tau}{1+P} \right)^{1/3} (r_m \tan \theta_m)^{2/3} 2^{-1/6} \quad [15b]$$

$$\omega_c = \left(\frac{\tau^2}{(1+P)^2 P} \right)^{1/3} 2^{-5/6} (\tan^2 \theta_m / r_m)^{2/3} \quad [15c]$$

where r_m refers to the mean radius of the fluid shell, $r_m = (r_i + r_o)/2$, and θ_m to the corresponding co-latitude, $\theta_m = \arcsin(r_m(1-\eta))$. The azimuthal wave number of the preferred mode is denoted by m_c and the corresponding angular velocity of the drift of the convection columns in the prograde direction is given by ω_c/m_c .

In [Figure 2](#) expressions [15a] and [15c] are compared with accurate numerical values which indicate that the general trend is well represented by

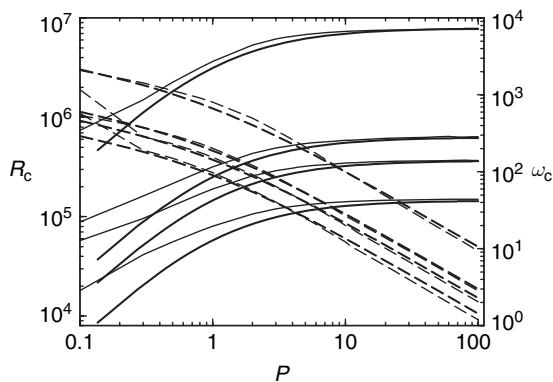


Figure 2 Critical Rayleigh number R_c (thin solid lines) and frequency ω_c (right ordinate, thin dashed lines) as a function of the Prandtl number P in the case $\eta = 0.4$ for the Coriolis numbers $\tau = 5 \times 10^3, 10^4, 1.5 \times 10^4$, and 10^5 (from bottom to top). The thick lines correspond to expressions [15a] and [15c].

expressions [15a] and [15c]. The same property holds for m_c . For a rigorous asymptotic analysis in the case $P=1$ including the radial dependence, we refer to Jones *et al.* (2000).

It is evident from **Figure 2** that the agreement between expressions [15] and the numerical values deteriorates as low values of P are approached. This behavior is caused by the fact that instead of the thermal Rossby wave mode the inertial mode of convection becomes preferred at onset for sufficiently low Prandtl numbers. It is characterized by convection cells attached to the equatorial region of the outer boundary not unlike the ‘banana cells’ seen in the narrow gap experiment of **Figure 3**. The equatorially attached convection does indeed represent an inertial wave modified by the effects of viscous dissipation and thermal buoyancy. An analytical description of this type of convection can thus be attained through the introduction of viscous friction and buoyancy as perturbations as has been done by Zhang (1994) and Busse and Simitev (2004) for stress-free and by Zhang (1995) for no-slip boundaries. According to Ardes *et al.* (1997), equatorially attached convection is preferred at onset for $\tau < \tau_l$ where τ_l increases in proportion to $P^{-1/2}$.

A third form of convection is realized in the polar regions of the shell which comprise the two fluid domains inside the tangent cylinder. Since gravity and rotation vectors are nearly parallel in these regions (unless values of η close to unity are used), convection resembles the kind realized in a horizontal layer heated from below and rotating

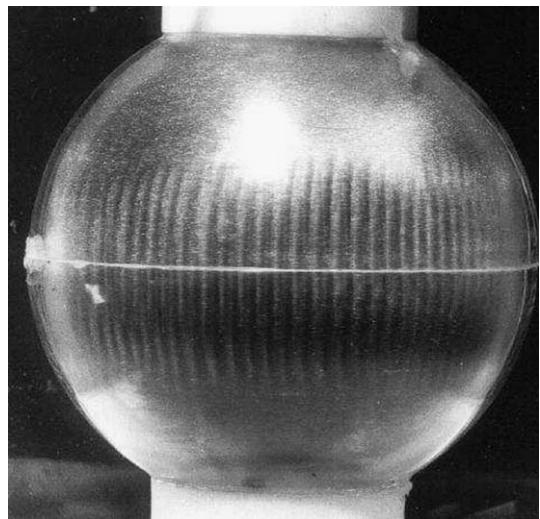


Figure 3 Banana cells in a thin rotating spherical fluid shell cooled from within. Convection driven by centrifugal buoyancy is made visible by a suspension of tiny flakes which become aligned with the shear. Adapted from Busse FH and Carrigan CR (1976) Laboratory simulation of thermal convection in rotating planet and stars. *Science* 191: 81–83.

about a vertical axis. Because the Coriolis force cannot be balanced by the pressure gradient in this case, the onset of convection is delayed to higher values of R where convection outside the tangent cylinder has reached already high amplitudes. In the case of $\eta = 0.4$ the onset of convection in the polar regions typically occurs at Rayleigh numbers which exceed the critical values R_c for onset of convection outside the tangent cylinder by a factor of the order 4. Except for the case of very low Prandtl numbers the retrograde differential rotation in the polar regions generated by convection outside the tangent cylinder tends to facilitate polar convection by reducing the rotational constraint. A tendency toward an alignment of convection rolls with the north–south direction (Busse and Cuong, 1977) can be noticed, but this property is superseded by instabilities of the Küppers-Lortz type (for an experimental demonstration, see Busse and Heikes (1980)) and by interactions with turbulent convection outside the tangent cylinder.

Typical features of low and high Prandtl number convection are illustrated in **Figure 4**. The columnar nature of convection does not vary much with P as is evident from the top two plots of the figure. At Prandtl numbers of the order unity or less, but not in the case of inertial convection, the convection columns tend to spiral away from the axis and

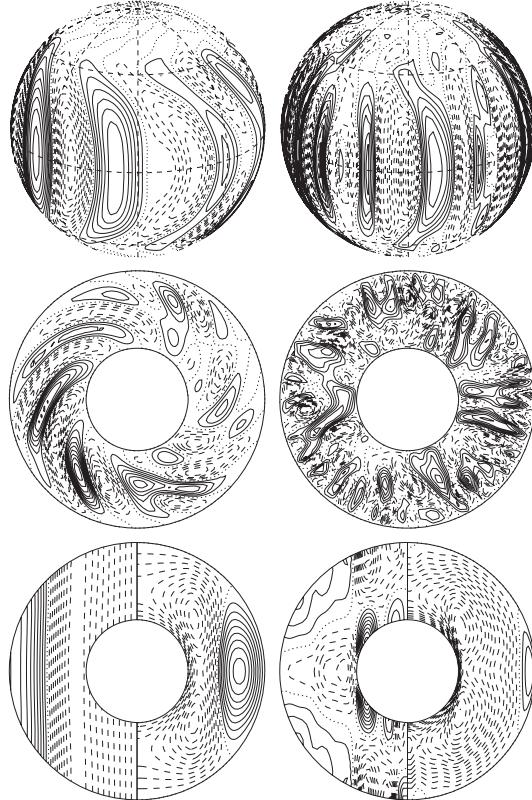


Figure 4 Convection in rotating spherical fluid shells in the cases $\tau = 10^4$, $R = 4 \times 10^5$, $P = 1$ (left column) and $\tau = 5 \times 10^3$, $R = 8 \times 10^5$, $P = 20$ (right column). Lines of constant u_r in the middle spherical surface, $r = r_i + 0.5$ are shown in the upper row. The plots of the middle row show streamlines, $r \partial\Phi/\partial\phi = \text{const.}$, in the equatorial plane. The lowermost plots indicate lines of constant mean azimuthal velocity \bar{u}_ϕ in the left halves and isotherms of $\bar{\Theta}$ in the right halves.

thereby create a Reynolds stress which drives a strong geostrophic differential rotation as shown in the bottom left plot of the figure. This differential rotation in turn promotes the tilt and a feedback loop is thus created. At high values of P the Reynolds stress becomes negligible and no significant tilt of the convection columns is apparent. In this case the differential rotation is generated in the form of a thermal wind caused by the latitudinal gradient of the axisymmetric component of Θ .

Among the properties of convection at finite amplitudes the heat transport is the most important one. Customarily its efficiency is measured by the Nusselt number which is defined as the heat transport in the presence of convection divided by the heat transport in the absence of motion. In the case of the spherical fluid shell two Nusselt numbers can be

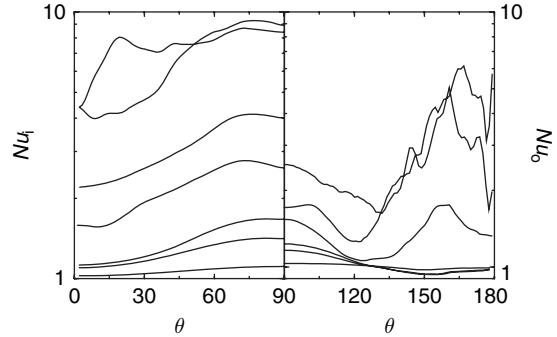


Figure 5 The time- and azimuthally-averaged local Nusselt numbers at the inner spherical boundary Nu_i (left plot) and at the outer spherical boundary Nu_o (right plot) as functions of the colatitude θ for $P = 0.5$, $\tau = 1.5 \times 10^4$ and $R = 4, 8, 10, 15, 20, 30, 25 \times 10^5$ (from bottom to top at 90°).

defined measuring the efficiency of convection at the inner and the outer boundary:

$$Nu_i = 1 - \frac{P d\bar{\Theta}}{r_i dr} \Big|_{r=r_i}, \quad Nu_o = 1 - \frac{P d\bar{\Theta}}{r_o dr} \Big|_{r=r_o} \quad [16]$$

where the double bar indicates the average over the spherical surface. In addition, local Nusselt numbers

$$Nu_i(\theta) = 1 - \frac{P d\bar{\Theta}}{r_i dr} \Big|_{r=r_i}, \quad Nu_o(\theta) = 1 - \frac{P d\bar{\Theta}}{r_o dr} \Big|_{r=r_o} \quad [17]$$

are of interest where only the azimuthal average is applied, as indicated by the single bar, instead of the average over the entire spherical surface. Examples of such measures of the dependence of the heat transport on latitude are shown in [Figure 5](#). The figure demonstrates that at low supercritical Rayleigh numbers the heat transport occurs primarily across the equatorial region, but as R increases the heat transport in the polar regions takes off and soon exceeds that at low latitudes. This process is especially evident at the outer boundary. In the polar regions, convection is better adjusted for carrying heat from the lower boundary to the upper one, and it is known from computations of the convective heat transport in horizontal layers rotating about a vertical axis that the value of Nu may exceed the value in a nonrotating layer at a given value of R in spite of the higher critical Rayleigh number in the former case ([Somerville and Lipps, 1973](#)).

Outside the tangent cylinder the convective heat transport encounters unfavorable conditions in that the cylindrical form of the convection eddies is not well adjusted to the spherical boundaries. This handicap is partly overcome through the onset of time dependence in the form of vacillations in which the

convection columns expand and contract in the radial direction or vary in amplitude.

At Prandtl numbers of the order unity and less another effect restricts the heat transport. The shear of the geostrophic differential rotation created by the Reynolds stresses of the convection columns severely inhibits the heat transport. To illustrate this effect we have plotted in **Figure 6** in addition to the Nusselt numbers the averages of the kinetic energy densities of the various components of the convection flow which are defined by

$$\bar{E}_p = \frac{1}{2} \langle |\nabla \times (\nabla \bar{\Phi} \times \mathbf{r})|^2 \rangle, \quad \bar{E}_t = \frac{1}{2} \langle |\nabla \bar{\Psi} \times \mathbf{r}|^2 \rangle \quad [18a]$$

$$\check{E}_p = \frac{1}{2} \langle |\nabla \times (\nabla \check{\Phi} \times \mathbf{r})|^2 \rangle, \quad \check{E}_t = \frac{1}{2} \langle |\nabla \check{\Psi} \times \mathbf{r}|^2 \rangle \quad [18b]$$

where the angular brackets indicate the average over the fluid shell and $\bar{\Phi}$ refers to the azimuthally averaged component of Φ , while $\check{\Phi}$ is defined by $\check{\Phi} = \Phi - \bar{\Phi}$. Analogous definitions hold for the energy densities of the magnetic field:

$$\bar{M}_p = \frac{1}{2} \langle |\nabla \times (\nabla \bar{b} \times \mathbf{r})|^2 \rangle, \quad \bar{M}_t = \frac{1}{2} \langle |\nabla \bar{g} \times \mathbf{r}|^2 \rangle \quad [19a]$$

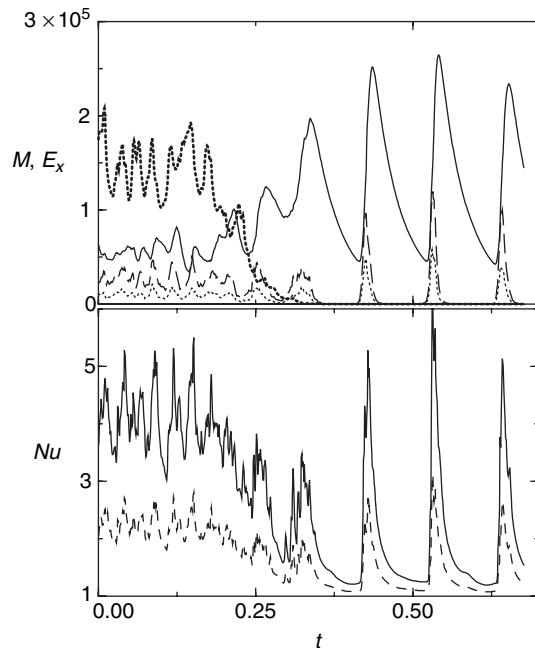


Figure 6 Decay of a dynamo for $P=0.5$, $\tau=1.5 \times 10^4$, $R=1.2 \times 10^6$, $P_m=0.5$. The total magnetic energy density M (thick dotted line, multiplied by a factor of 5), the kinetic energy densities \bar{E}_t (solid line), \check{E}_p (dotted line), \check{E}_t (dashed line), and Nusselt numbers Nu_i (solid line) at the inner and Nu_o (dashed line) at the outer spherical boundary (lower plot) are plotted as functions of time t .

$$\check{M}_p = \frac{1}{2} \langle |\nabla \times (\nabla \check{\Phi} \times \mathbf{r})|^2 \rangle, \quad \check{M}_t = \frac{1}{2} \langle |\nabla \check{\Psi} \times \mathbf{r}|^2 \rangle \quad [19b]$$

The total magnetic energy density M averaged over the fluid shell is thus given by $M = \bar{M}_p + \bar{M}_t + \check{M}_p + \check{M}_t$.

Figure 6 is instructive in that it demonstrates both, convection in the presence and in the absence of its dynamo generated magnetic field. As is evident from the right part of **Figure 6** with $M \sim 0$, relaxation oscillations occur in which convection sets in nearly periodically for short episodes once the energy \bar{E}_t of the differential rotation has decayed to a sufficiently low amplitude. But as soon as convection grows in amplitude, the differential rotation grows as well and shears off the convection columns. After convection has stopped the differential rotation decays on the viscous timescale until the process repeats itself.

As long as the magnetic field is present \bar{E}_t is suppressed owing to the action of the Lorentz force and high Nusselt numbers are obtained. The dynamo-generated magnetic field thus acts in a fashion quite different from that of a homogeneous field which typically counteracts the effects of rotation and tends to minimize the critical value of the Rayleigh number when the Elsasser number

$$\Lambda = \frac{2MP_m}{\tau} \quad [20]$$

assumes the value 1 in the case of a plane layer (Chandrasekhar, 1961) or values of the same order in the case of a sphere (Fearn, 1979) or in the related annulus problem (Busse, 1983).

10.08.5 Convection-Driven Dynamos

It appears that dynamos are generated by convection in rotating spherical shells for all parameters values as long as the magnetic Reynolds number is of the order 50 or higher and the fluid is not too turbulent. Since R_m can be defined by $P_m \sqrt{2E}$ where the kinetic energy density, $E = \bar{E}_p + \bar{E}_t + \check{E}_p + \check{E}_t$ increases with R , increasing values of the Rayleigh number are required for dynamos as P_m decreases. In planetary cores the latter parameter may assume values as low as 10^{-5} or 10^{-6} , but numerical simulation has achieved so far only values somewhat below 10^{-1} . Numerical simulations of the geodynamo are discussed in Chapter 8.08. The trend of increasing R with decreasing P_m is evident in **Figure 7** where results are shown for two different values of τ and

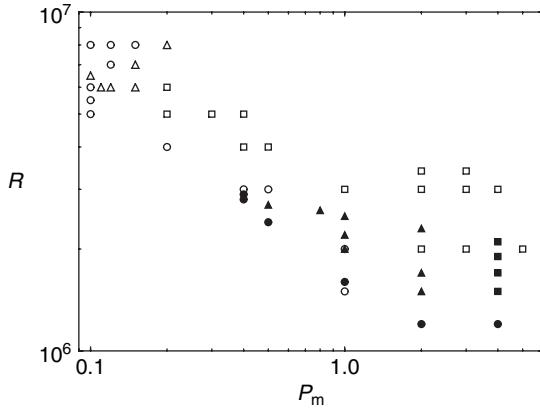


Figure 7 Convection-driven dynamos as a function of the Rayleigh number R and the magnetic Prandtl number P_m for $P = 0.1$, $\tau = 10^5$ (empty symbols) and for $P = 1$, $\tau = 3 \times 10^4$ (filled symbols). The symbols indicate chaotic dipolar (squares), hemispherical (triangles), and decaying dynamos (circles).

P . It is also evident from the results in the upper left corner of the figure that an increasing R may be detrimental for dynamo action. This is a typical property of marginal dynamos at the boundary of dynamos in the parameter space (Christensen *et al.*, 1999; Simitev and Busse, 2005). Since convection at the relevant values of R is chaotic, this property also holds for the generated magnetic field. The magnetic energy must be finite near the boundary of dynamos in the parameter space because a sufficient amplitude of the fluctuating components of the convection flow can be obtained only when the magnetic field is strong enough to suppress the differential rotation. This property appears to hold even at high Prandtl numbers where differential rotation occurs only in the form of a relatively weak thermal wind.

Even in its chaotic state, convection exhibits a strong symmetry with respect to the equatorial plane at high values of τ , at least as long as convection in the polar regions is still weak. Because of this symmetry magnetic fields can be generated either with the same symmetry as the convection flow, that is, b and g are antisymmetric with respect to equatorial plane in which case one speaks of a quadrupolar dynamo, or the magnetic field exhibits the opposite symmetry with symmetric functions b and g in which case one speaks of a dipolar dynamo. Of special interest are hemispherical dynamos (Grote and Busse, 2001) in which case the fields of dipolar and quadrupolar symmetry have nearly the same amplitude such that they cancel each other either in

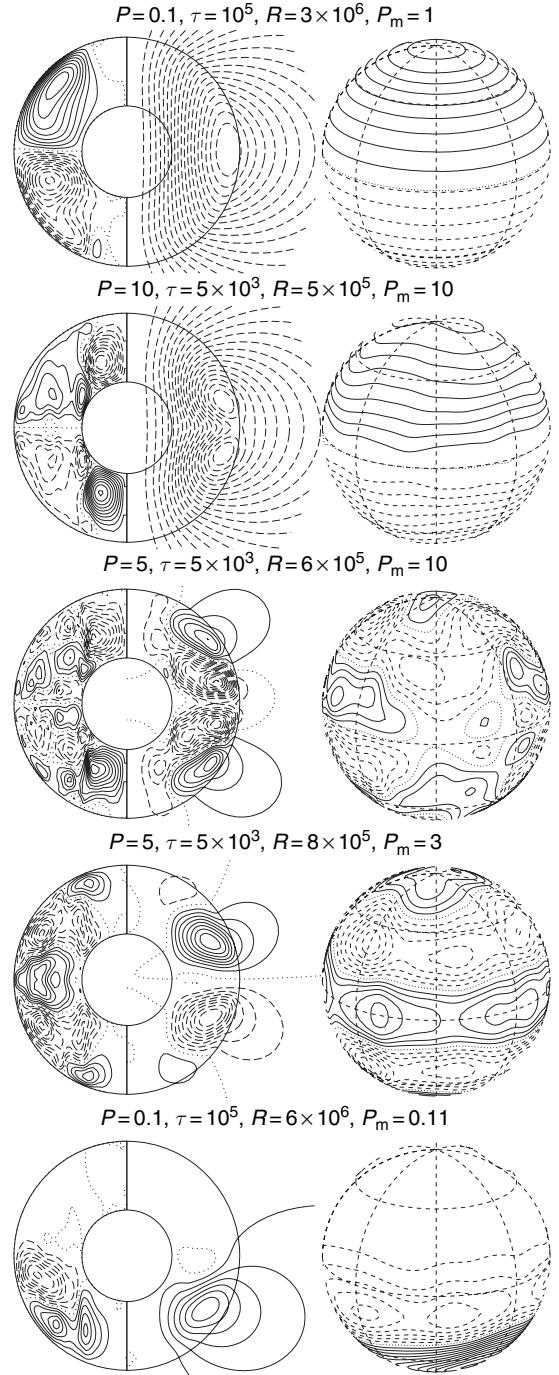


Figure 8 Various types of dynamo symmetry. The left column shows lines of $B_\phi = \text{const.}$ (left half) and $r \sin \theta \partial_\theta h = \text{const.}$ (right half). The right column shows surfaces of $B_r = \text{const.}$ at $r = r_o + 0.7$.

the Northern or the Southern Hemisphere. Examples of typical dynamos with different symmetries are shown in **Figure 8**. Here it is also evident that low

P and high P dynamos differ in the structure of their mean toroidal fields. While their mean poloidal dipolar fields exhibit hardly any difference, the high P dynamo is characterized by strong polar azimuthal flux tubes which are missing in the low P case. The reason for this difference is that the radial gradient of the differential rotation in the polar regions is much stronger for high P than for low P .

As the Rayleigh number increases and polar convection becomes stronger, the convection flow loses some of its equatorial symmetry and the magnetic field can no longer be easily classified. Usually, the dipolar component becomes more dominant in cases of dynamos which started as either quadrupolar or hemispherical dynamos at lower values of R .

It should be emphasized that quadrupolar and hemispherical dynamos are usually oscillatory in that new azimuthal flux tubes of alternating sign emerge in the equatorial plane, move toward higher latitudes, while old flux tubes are dissipated in the polar region. There also exist dipolar dynamos which exhibit the same type of oscillations. They are typically found in the parameter space near the region where hemispherical dynamos occur. To illustrate

the dynamo oscillations additional plots separated at equal distances in time before the corresponding plot in the lower three rows of [Figure 8](#) are shown in [Figure 9](#). The four plots in each case cover approximately half a period of oscillation. It must be realized, of course, that the oscillations are not strictly periodic since they occur in a turbulent system. In this respect they resemble the solar cycle with its 22-year period. Because the solar dynamo is believed to operate at the bottom of the solar convection, the propagation of the dynamo wave is toward low latitudes on the Sun. A remarkable feature of the dipole oscillation in [Figure 9](#) is that the polar flux tubes hardly change in time. The oscillation appears to be confined to the region outside the tangent cylinder. At a lower value of P_m the strong polar flux tubes even inhibit the oscillation of the mean poloidal field as shown in [Figure 10](#). This case has been called the ‘invisibly’ oscillating dynamo since at some distance from the boundary of the spherical fluid shell the oscillation of the dynamo can hardly be noticed.

The relatively simple magnetic fields displayed in [Figures 8–10](#) should not be regarded as representative for high values of the Rayleigh number when

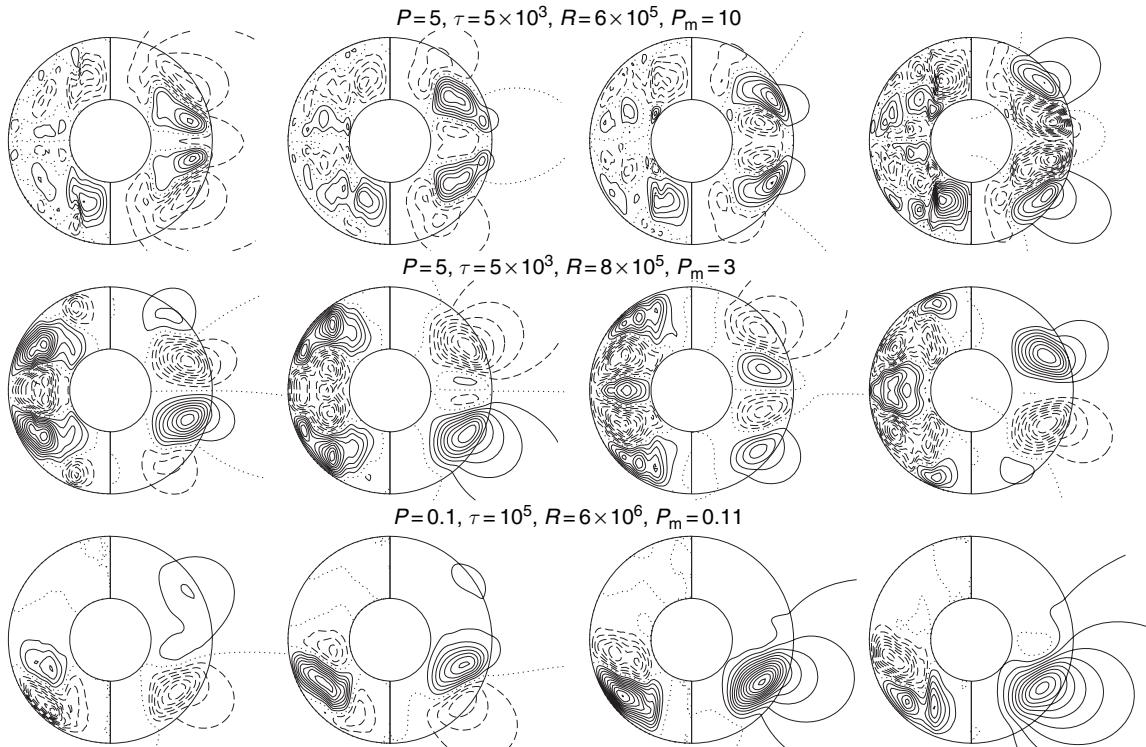


Figure 9 Plots for the last three cases of [Figure 8](#) at the times $n \times \Delta t, n = 3, 2, 1, 0$ (left to right) before the times of [Figure 8](#) with $\Delta t = 0.035, 0.04, 0.0025$ (from top to bottom).

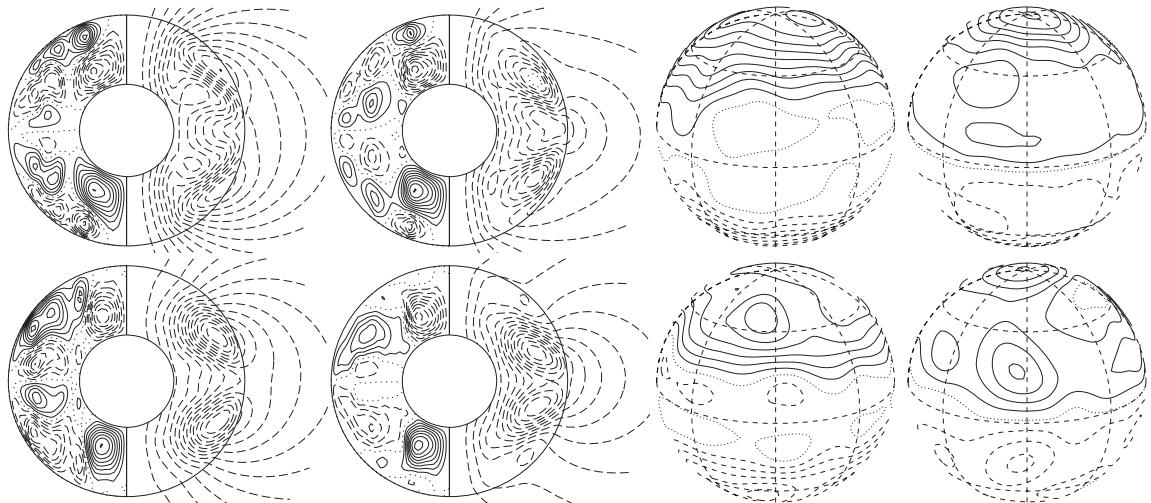


Figure 10 An ‘invisibly’ oscillating dynamo with $\Delta t = 0.04$ in the case of $P = P_m = 5$, $\tau = 5 \times 10^3$ and $R = 6 \times 10^5$. The plots to the left show lines of constant B_ϕ in their left halves and meridional field lines, $r \sin \theta \partial_\theta \bar{h} = \text{const.}$, in their right halves. The plots to the right exhibits lines of constant B_r on the surface $r = r_o + 0.4$. The plots follow in clockwise sense starting with the upper left one such that they complete approximately a full cycle.

R exceeds its critical value R_c by an order of magnitude or more. In that case the equatorial symmetry has nearly disappeared owing to the dominant convection in the polar regions and nonaxisymmetric components of the field tend to exceed axisymmetric ones as shown, for example, in [Figure 11](#). The strong growth with increasing R_m of the nonaxisymmetric magnetic flux is also evident in [Figure 12](#) when energies corresponding to different values of P_m are compared.

An important question is the average strength of the magnetic field generated by the dynamo process in dependence on the parameters. Two concepts are often used in attempts to answer this question. In astrophysical situations, the equipartition between magnetic and kinetic energy is used as a guide. While such a balance may hold locally as, for example, in the neighborhood of sunspots, it is not likely to be applicable to global planetary magnetic fields.

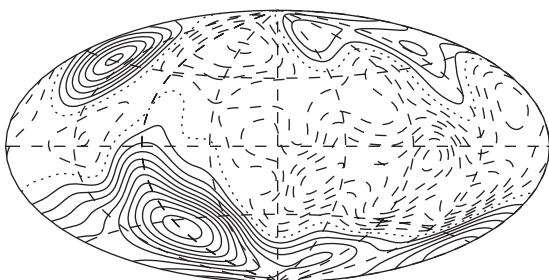


Figure 11 Lines of constant B_r at $r = r_o + 1$ for the dynamo with $R = 2 \times 10^6$, $\tau = 5 \times 10^3$, $P = P_m = 1$.

In the Earth’s core, for example, the magnetic energy density exceeds the kinetic one by a factor of the order of 10^3 . The second concept is based on the property that an Elsasser number Λ of the order unity appears to correspond to optimal conditions for convection in rotating systems in the presence of an applied nearly uniform magnetic field. While this idea may be useful as a first rough estimate, the results of numerical dynamo simulations do not support this concept very well (*see* Chapters 8.08, 8.11). The examples shown in [Figure 12](#) exhibit Λ -values differing by an order of magnitude and even larger variations have been reported by [Simitev and Busse \(2005\)](#) and [Christensen and Aubert \(2006\)](#).

Another important feature demonstrated in [Figure 12](#) is the change in the structure of the magnetic field with increasing Prandtl number. While for $P = 5$ and lower values the mean poloidal field is small in comparison with the fluctuating components, this situation changes dramatically at about $P = 8$ for $\tau = 5 \times 10^3$ such that for higher P the mean poloidal field becomes dominant. Usually, this field is dipolar. This change is associated with the transition from the geostrophic differential rotation generated by Reynolds stress to the thermal wind-type differential rotation caused by a latitudinal temperature gradient. While the magnetic energy M may exceed the total kinetic energy E by orders of magnitude in particular for high Prandtl numbers, Ohmic dissipation is usually found to be at most comparable

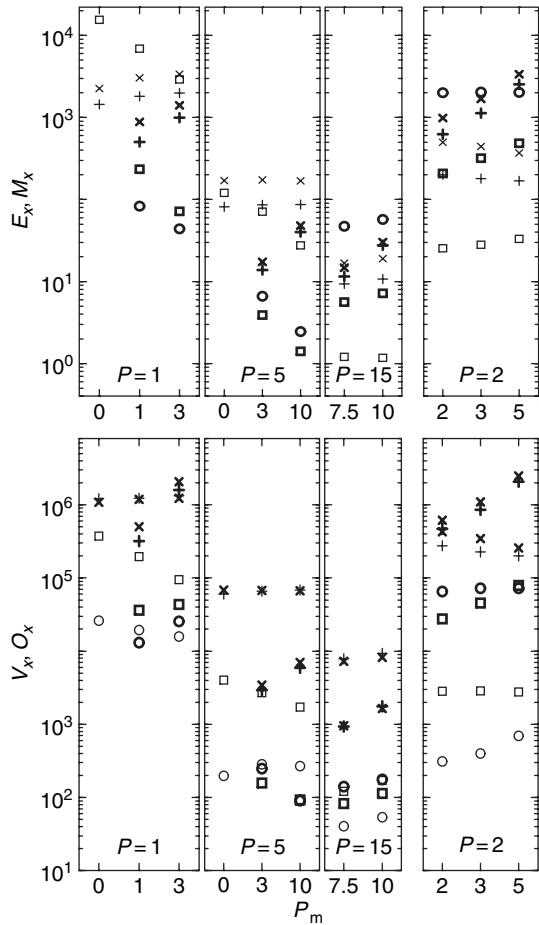


Figure 12 Kinetic E_x and magnetic M_x energy densities (upper part) and viscous V_x and Ohmic O_x dissipations (lower part) as functions of P_m for convection-driven dynamos for $\tau = 5 \times 10^3$, $R = 10^6$ (left three columns), $\tau = 3 \times 10^4$, $R = 3.5 \times 10^6$ (right column) and Prandtl number as indicated in the boxes. The highest values of the Elsasser number for the cases $P=1, 5, 15, 2$ are $\Lambda = 3.02, 0.37, 0.49$, and 1.12 , respectively. The components $X_p, X_t, \dot{X}_p, \dot{X}_t$ (where $X = E, M, V, O$) are represented by circles, squares, plus-signs, and crosses, respectively. Kinetic energy densities and viscous dissipations are shown with light symbols, magnetic energy densities, and Ohmic dissipations are shown with heavy symbols. The magnetic Prandtl numbers $P_m = 0$ correspond to nonmagnetic convection.

to viscous dissipation in numerical simulations. But this may be due to the limited numerically accessible part of the parameter space.

The brief introduction of this section to convection-driven dynamos in rotating spherical fluid shells can only give a vague impression of the potential of numerical simulations for the understanding of planetary magnetism. Many more examples of such simulations can be found in the literature. Usually,

they have been motivated by applications to the geodynamo and $P = 1$ is assumed in most cases for simplicity. For some recent systematic investigations, we refer to Christensen *et al.* (1999), Grote *et al.* (2000), Jones (2000), Grote and Busse (2001), Kutzner and Christensen (2000, 2002), Busse *et al.* (2003), Simitev and Busse (2005) and other papers referred to therein.

10.08.6 Applications to Planetary Dynamos

10.08.6.1 General Considerations

In many respects it is too early to model the dynamo process in particular planets. The numerical simulations of the kind discussed in the preceding sections are still rather removed from the parameter regime relevant to planetary interiors. Only most recently, attempts have been made to extrapolate results to high Rayleigh numbers and high values of τ (Christensen and Aubert, 2006). Moreover, only the most important physical parameters have been taken into account and relevant properties such as compressibility and other deviations from the Boussinesq approximation have been considered only in special models applied to the Earth's core (Glatzmaier and Roberts, 1996) or to the Sun (Brun *et al.*, 2004). On the other hand, many essential parameter values of planetary cores are not sufficiently well known to provide a basis for the development of specific dynamo models. Much of the future progress of the field will thus depend on the mutual constraints derived from observational evidence and from theoretical conclusions in order to arrive at a better understanding of the workings of planetary dynamos.

In Table 2 some properties related to planetary magnetism have been listed. The dipole moments of the planets have been given as multiples of $10^{-4} T \times R_p^3$, where R_p denotes the mean radius of the planet or satellite. Thus, the numerical value indicates the field strength in Gauss in the equatorial plane of the dipole at the distance R_p from its center. We have included Mars in the table, although it does have an active dynamo at the present time. The strong magnetization of the Martian crust indicates, however, that a strong field dynamo must have operated in the early history of the planet.

The most important question of the theory of planetary magnetism is the dependence of the observed field strength on the properties of the planet. We have already discussed in the historical introduction various

Table 2 Planetary parameter values (after Stevenson (2003), Nellis et al. (1988), Connerney (1993), and various websites)

Planet (satellite)	Planetary radius $R_p(10^6 \text{ m})$	Dipole moment ($10^{-4} \text{ T} \times R_p^3$)	Core radius (10^6 m)	Angular rate of rot. $\Omega(10^{-5} \text{ s}^{-1})$	Magn. diffus. ($\text{m}^2 \text{ s}^{-1}$)
Mercury	2.439	0.0025	1.9	0.124	~ 2
Earth	6.371	0.31	3.48	7.29	2
Mars	3.389	3^b	~ 1.5	7.08	~ 2
Jupiter	69.95	4.3	60 ^a	17.6	100 ^a
Saturn	58.30	0.21	35 ^a	16.2	100 ^a
Uranus	25.36	0.23	17 ^a	10.1	100 ^a
Neptune	24.62	0.14	18 ^a	10.8	100 ^a
Ganymede	2.63	0.0072	0.7	1.02	4

^aSince the giant planets do not possess a well-defined boundary of a highly conducting core, values of the most likely region of dynamo activity are given.

^bRemnant magnetism of the Martian crust appears to require an ancient dynamo with a field strength of at least 10 times the Earth's magnetic field.

proposals for such dependences based on *ad hoc* assumptions, but these have been abandoned by and large. Only the concept of an Elsasser number Λ of the order unity is still frequently used. Stevenson (2003) shows that a value $\Lambda = 0.3$ fits most of the planets with a global magnetic field quite well, but in cases such as Mercury and Saturn only a value of the order 10^{-2} can be estimated for Λ . For the observed properties of the magnetic fields of the planets other than the Earth the reader is referred to Chapter 10.07. In order to save the concept, different dynamo regimes must be assumed. Equilibration balances for weaker magnetic fields have been proposed as, for instance, in the case when the strength of convection is not sufficient to attain the $\Lambda \sim 1$ balance. For details, see Stevenson (1984).

An important balance often invoked in discussion of planetary dynamos (see, e.g., Stevenson (1979) and Jones (2003)) is the MAC-balance (Braginsky, 1967) where it is assumed that Lorentz force, buoyancy force, and pressure gradient are all of the same order as the Coriolis force, while viscous friction and the momentum advection term are regarded as negligible in the equations of motion. The neglection of the latter term is well justified for high Prandtl numbers, but it is doubtful for values of P of the order unity or less. In spite of their smallness in comparison with the Coriolis force, the divergence of the Reynolds stress can generate the most easily excitable mode of a rotating fluid, namely the geostrophic differential rotation which cannot be driven by the Coriolis force, the pressure gradient, or buoyancy. Of course, an excitation by the Lorentz force is possible in principle. The latter force, however, usually inhibits the geostrophic differential rotation. Since the mean azimuthal component of the magnetic field is typically created through the shear of the differential

rotation, the Lorentz force opposes the latter according to Lenz' rule.

Christensen and Aubert (2006) have recently introduced a concept in which the equilibrium strength of planetary magnetism does no longer depend on the rate of rotation Ω , but instead depends only on the available power for driving the dynamo. They find that scaling laws can be obtained once Rayleigh number, Nusselt number, and dimensionless buoyancy flux \mathcal{Q} have been defined through quantities that no longer involve molecular diffusivities. This is a surprising result since it requires the presence of rotation and cannot be achieved in a nonrotating system. The final estimate obtained for the strength of the magnetic field in the dynamo region fits the cases of the Earth and Jupiter quite well, but yields the result that Mercury's field could not be generated by a buoyancy-driven dynamo since the magnetic Reynolds number would be too small. In the cases of the outer planets relevant parameters are not sufficiently well known to draw definitive conclusions. The negative result in the case of Mercury may be misleading, however, since the observed field may not be representative for the dynamo generated field inside the planet as will be discussed in Section 10.08.6.2.

Finally, we like to draw attention to what has been called Stevenson's paradox, namely that a too high electrical conductivity is detrimental for dynamo action in planetary cores. According to the Wiedemann–Franz law of condensed matter physics, the thermal conductivity k of a metal is proportional to its electrical conductivity times the temperature, $k = \sigma TL$, where L has approximately the same value for all metals. This is due to the fact that in metals electrons dominate both, the transports of heat and of

charge. As a consequence there is an upper bound on the electrical conductivity for which a dynamo driven by thermal convection is possible because a high value of σ implies a low-temperature change with increasing depth needed to carry the heat from the interior of the planetary core. But as soon as the temperature change with pressure falls below its isentropic value, thermal convection disappears and a stably stratified environment is obtained which makes it difficult to drive a dynamo by other types of motions such as compositional convection. Ignoring the later possibility, Stevenson (2003) has sketched the diagram shown in Figure 13 where the marginal nature of dynamos in terrestrial as well as in the ‘icy’ planets is indicated.

Related to Stevenson’s paradox is the fact that regions of high electrical conductivity are not the most likely places where dynamos operate in the major planets. Magnetic field generation in regions of metallic conductivity in Jupiter and Saturn would lead to an Ohmic dissipation exceeding the luminosity of these planets as has recently been emphasized by Liu *et al.* (2007). It thus appears likely that the dynamos of these planets operate in regions where hydrogen is still a semiconductor.

In the following brief characterizations of the magnetisms of various planets and moons are given. In terrestrial planets and satellites, dynamos are possible in the liquid parts of their iron cores. Such cores

always include light elements which depress the temperature of freezing and can give rise to compositional buoyancy in the presence of a solidifying inner part of the core. In Jupiter and especially in Saturn convection may also be driven in part by compositional buoyancy since hydrogen and helium are immiscible in certain regions and helium drops may rain out. In numerical dynamo models, the different sources of buoyancy are usually not distinguished, however.

10.08.6.2 Mercury

It has already become apparent that the origin of Mercury’s magnetism is especially enigmatic. Besides the hypothesis of an active dynamo, thermoelectric currents and remnant magnetism of the crust compete as explanations. The latter possibility had been discarded for a long time since Runcorn’s theorem states that a homogeneous spherical shell cannot be magnetized by an interior magnetic field in such a way that a dipolar field can be observed from the outside after the interior source has been removed (Runcorn, 1975). Aharonson *et al.* (2004) have shown, however, that plausible inhomogeneities of the crust could possibly explain a remnant magnetic field created by an ancient dynamo.

The idea that a convection-driven dynamo operates in a fluid outer core of Mercury has received some support from observations of Mercury’s librations (Margot *et al.*, 2004) and from numerical simulations (Stanley *et al.*, 2005). The amplitudes of the small periodic variations (librations) of Mercury’s rotation caused by the gravitational pull of the Sun on the nonaxisymmetric mass distribution of Mercury appear to be too large for a fully solidified planet. This suggests that the solid inner core is separated from the mantle by a fluid outer core in which convection flows may occur. The numerical simulations by Stanley *et al.* (2005) of dynamos in a thin shell outer liquid core have demonstrated that the dipole strength measured from the outside may not be representative for the strength of the magnetic field at the core–mantle boundary and the criteria mentioned in Section 10.08.6.1 may be satisfied after all.

Another explanation for the low strength of the observed Mercurian magnetic field has been given by Christensen (2006). He advocates a stably stratified outer part of the liquid outer core which effectively damps the field generated by convection in the inner part of the liquid outer core. In this way it can be demonstrated that Mercury’s dynamo may fit the

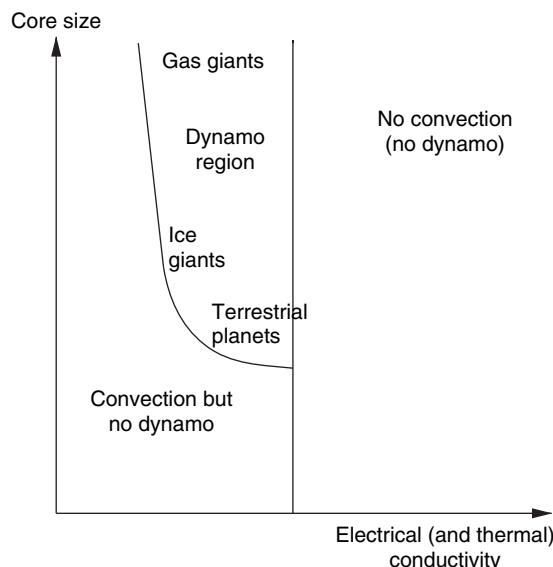


Figure 13 Sketch of Stevenson’s paradox. After Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11.

scaling relationship of Christensen and Aubert (2006). The question, however, whether there is a sufficiently strong source of buoyancy to sustain convection and the dynamo over the age of the planet has not yet been answered satisfactorily.

10.08.6.3 Venus

Space probes have clearly shown that there is no dynamo operating presently in Venus. This result is surprising since Venus is very similar to the Earth in many other aspects. That the period of rotation is only 1/243 of a day should not matter much since the Coriolis parameter τ is still huge. More important is the fact that the cooling of the planet is less efficient because of the apparent absence of mantle convection and plate tectonics. A solid inner core may not have yet started to grow in Venus and vigorous convection does not seem to occur in the core, at least not at the present time.

10.08.6.4 Mars and Moon

Neither Moon nor Mars have an active dynamo, but the magnetized rocks in their crusts suggest that in their early history about 4 billion years ago, dynamos may have been operating in the iron cores of these bodies. The small lunar core with a radius of the order of 350 km and the age of the magnetized lunar rocks put severe constraints on a possible dynamo origin of lunar magnetism as discussed by Stegman *et al.* (2003). In their model these authors try to accommodate in particular the apparent sudden onset of a lunar dynamo at a time 3.5 Ga before present. Because of its larger iron core the possibility of a temporarily operating Martian dynamo is more likely, but the strong magnetization of parts of the Martian crust requires a field of at least 10 times the strength of the present geomagnetic field. For a recent review of the implications of Martian magnetism for the evolution of the planet, we refer to Stevenson (2001).

10.08.6.5 Jupiter

The Jovian magnetic field is rather similar to the geomagnetic one sharing with it a dipole axis that is inclined by an angle of the order of 10° with respect to the axis of rotation. That the higher harmonics of the field are relatively stronger than those in the case of the Earth suggest that the Jovian dynamo is driven at a more shallow layer than the Earth's outer core. This assumption is in accordance with the transition to

metallic hydrogen which is expected to occur at a pressure of 140 GPa corresponding to a depth of about 15% of the Jovian radius (Nellis *et al.*, 1996). The transition is not a sharp one as suggested by earlier models of Jupiter's interior, but a gradual one leading to a decrease of the magnetic diffusivity down to about $4 \text{ m}^2 \text{ s}^{-1}$. Since this value yields a rather high value for the Elsasser number Λ , Stevenson (2003) has argued that the dynamo is most active above the region of highest conductivity. Jones (2003), on the other hand, accepts the value 20 for Λ , but must admit a high magnetic Reynolds number of the order 10^4 at which an effective dynamo may no longer be possible. A detailed model of a convection-driven Jovian dynamo should include effects of compressibility and depth-dependent electrical conductivity, but none has yet been published.

10.08.6.6 Saturn

Saturn's magnetic field with a strength of 0.2 mT near the poles is much weaker than that of Jupiter which must be attributed primarily to its deeper origin in the planet. The transition pressure of 140 GPa is reached only at a depth of about 40% of the planetary radius. A property that has received much attention is that the Saturnian magnetic field is nearly axisymmetric with respect to the axis of rotation. This property cannot be interpreted as a contradiction to Cowling's (1934) theorem since it has long been shown that fields with arbitrarily small deviations from axisymmetry can be generated by the dynamo process (Braginsky, 1976). Stevenson (1982) gave an explanation for the almost axisymmetric Saturnian field by demonstrating that the differential rotation in the stably stratified, but still electrically conducting gas shells above the dynamo region, will tend to shear off all nonaxisymmetric components of the field. The problem may be a bit more intricate as has been pointed out by Love (2000) who showed that sometimes fields with solely nonaxisymmetric components are found in such situations.

10.08.6.7 Uranus and Neptune

A metallic liquid probably does not exist in Uranus and Neptune whose interiors consist mostly of ice and rocks. Because of the pressure dissociation of water it is believed that deeper regions of these planets are characterized by an ionic conductivity which is lower than typical metallic conductivities by between one and two orders of magnitude. Recent

laboratory measurements (Lee *et al.*, 2006) suggest, however, that water is semiconducting at pressures of 100 GPa and relevant temperatures of the order of 4000 K. The mixture of water, methane, ammonia, and other ices is sufficiently fluid that convection can occur and that a dynamo is thus possible. The fact that the observed magnetic field do not show an alignment with the axis of planetary rotation and that quadrupolar and octupolar components are comparable to the dipolar components of the fields has been attributed to a dynamo operating in a thin shell (Ruzmaikin and Starchenko, 1991; Stanley and Bloxham, 2006). This seems to be hardly necessary, however, since even convection-driven dynamos in thick shells often exhibit such fields when small-scale components dominate as in the case of high Rayleigh number and low Prandtl numbers (see, for example, Figure 11 which shows a magnetic field rather similar to that of Uranus as displayed in the paper of Connerney (1993)). A special difficulty for a convection-driven dynamo in the interior of Uranus is caused by its low emission of heat. Holme and Bloxham (1996) suggest that a typical dynamo would involve more Ohmic dissipation than corresponds to the heat flux emitted from the interior of the planet.

10.08.6.8 Ganymede and Other Satellites

It came as a great surprise when the measurements of the Galileo spacecraft indicated that Jupiter's moon Ganymede possesses a global magnetic field for which an active dynamo inside the satellite seems to be the only realistic explanation (Kivelson *et al.*, 1996), although the possibility of a remnant magnetism cannot be easily excluded (Crary and Begenal, 1998). It must be kept in mind that Ganymede is the largest satellite in the solar system which exceeds even Mercury in size. Nevertheless, the estimated radius of its iron core is only about 660 km and it is hard to believe that it could still be partly molten unless it contains a lot of radioactive potassium 40 or Ganymede was captured into a resonance in its more recent history (Showman *et al.*, 1997). The fact that Ganymede's magnetic moment is nearly aligned with the ambient Jovian magnetic has led to the suggestion that the ambient field could aid significantly Ganymede's dynamo. Sarson *et al.* (1997) have investigated this question with the result that the ambient field is too weak to exert much influence. On the other hand, in the case of Io which is much closer to Jupiter, the interaction between the ambient field and

the liquid iron core can explain the observed structure of the magnetic field without the assumption of an active dynamo.

There are no other satellites in the solar system where an active dynamo must be suspected. The substantial magnetization of many iron meteorites suggest, however, that several differentiated protoplanets have had dynamos in the early days of the solar system.

10.08.7 Concluding Remarks

It is apparent from the above discussions that dynamo theory does not yet have much specific information to contribute to the interpretation of the observed magnetic properties of planets and satellites in the solar system. Even possible interactions between thermal and compositional buoyancies have not yet been taken fully into account (Glatzmaier and Roberts, 1996; Busse, 2002b). Some typical properties are already apparent, however, as for instance:

- Dynamos that are dominated by a nearly axial dipole and exhibit a magnetic energy that exceeds the kinetic one by orders of magnitude as in the case of the Earth are typical for high effective Prandtl numbers as must be expected when convection is primarily driven by compositional buoyancy. The rather Earth-like appearance of the dynamo of Glatzmaier and Roberts (1995) is in part due to the high value of P used in their computational model.
- Dynamos exhibiting strong higher harmonics are more likely driven by thermal convection corresponding to Prandtl numbers of the order unity or less.
- Dynamo oscillations are a likely phenomenon in the presence of a sufficiently strong differential rotation. They may not always be apparent in the poloidal magnetic field seen at a distance from the dynamo region.
- Considerations based on the effects on convection of an imposed homogeneous magnetic field cannot directly be applied to the case of convection-driven dynamos.

More details on the parameter dependence of dynamos will certainly emerge in the future as the increasingly available computer capacity will allow extensions of the parameter space accessible to computer simulations. Space probes such as MESSENGER in the case of Mercury will provide much needed detailed information on planetary

magnetic fields. Eventually, we may learn more about their variation in time which is one of the most interesting properties of planetary magnetic fields.

As in the case of stellar magnetism where the study of star spots and stellar magnetic cycles is contributing to the understanding of solar magnetism, it may eventually become possible to learn about the magnetism of extrasolar planets and apply this knowledge for an improved understanding of solar system dynamos. Undoubtedly, the field of planetary magnetism will continue to be an exciting one!

References

- Aharonson O, Zuber MT, and Solomon SC (2004) Crustal remanence in an internally magnetized non-uniform shell: A possible source for Mercury's magnetic field? *Earth and Planetary Science Letters* 218: 261–268.
- Ardes M, Busse FH, and Wicht J (1997) Thermal convection in rotating spherical shells. *Physics of the Earth and Planetary Interiors* 99: 55–67.
- Backus G (1958) A class of self-sustaining dissipative spherical dynamos. *Annals of Physics* 4: 372–447.
- Braginsky SI (1967) Magnetic waves in the Earth's core. *Geomagnetism and Aeronomy* 7: 851–859.
- Braginsky SI (1976) On the nearly axially-symmetrical model of the hydromagnetic dynamo of the Earth. *Physics of the Earth and Planetary Interiors* 11: 191–199.
- Brun AS, Miesch MS, and Toomre J (2004) Global-scale turbulent convection and magnetic dynamo action in the solar envelope. *Astrophysical Journal* 614: 1073–1098.
- Bullard EC (1949) The magnetic field within the Earth. *Proceedings of the Royal Society of London A* 197: 433–463.
- Burke BF and Franklin KL (1955) Observations of variable radio source associated with the planet Jupiter. *Journal of Geophysical Research* 60: 213–217.
- Busse FH (1970) Thermal instabilities in rapidly rotating systems. *Journal of Fluid Mechanics* 44: 441–460.
- Busse FH (1973) Generation of magnetic fields by convection. *Journal of Fluid Mechanics* 44: 441–460.
- Busse FH (1976) Generation of planetary magnetism by convection. *Physics of the Earth and Planetary Interiors* 12: 350–358.
- Busse FH (1983) Recent developments in the dynamo theory of planetary magnetism. *Annual Review of Earth and Planetary Sciences* 11: 241–268.
- Busse FH (2002a) Convection flows in rapidly rotating spheres and their dynamo action. *Physics of Fluids* 14: 1301–1314.
- Busse FH (2002b) Is low Rayleigh number convection possible in the Earth's core? *Geophysical Research Letters* 29: GLO149597.
- Busse FH and Carrigan CR (1976) Laboratory simulation of thermal convection in rotating planets and stars. *Science* 191: 81–83.
- Busse FH and Cuong PG (1977) Convection in rapidly rotating spherical fluid shells. *Geophysical and Astrophysical Fluid Dynamics* 8: 17–41.
- Busse FH, Grote E, and Simitev R (2003) Convection in rotating spherical shells and its dynamo action. In: Jones CA, Soward AM, and Zhang K (eds.) *Earth's Core and Lower Mantle*, pp. 130–152. London: Taylor and Francis.
- Busse FH and Heikes KE (1980) Convection in a rotating layer: A simple case of turbulence. *Science* 208: 173–175.
- Busse FH and Simitev R (2004) Inertial convection in rotating fluid spheres. *Journal of Fluid Mechanics* 498: 23–30.
- Christensen UR (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 44: 1056–1058.
- Christensen UR and Aubert J (2006) Scaling laws for dynamos in rotating spherical shells. *Geophysical Journal International* 166: 97–114.
- Christensen UR, Aubert J, Cardin P, et al. (2001) A numerical dynamo benchmark. *Physics of the Earth and Planetary Interiors* 128: 25–34.
- Christensen U, Olson P, and Glatzmaier GA (1999) Numerical modeling of the geodynamo: A systematic parameter study. *Geophysical Journal International* 138: 393–409.
- Chandrasekhar S (1961) *Hydrodynamic and Hydromagnetic Stability*. Oxford: Clarendon Press.
- Childress S and Soward AM (1972) Convection-driven hydrodynamic dynamo. *Physical Review Letters* 29: 837–839.
- Connerney JEP (1993) Magnetic fields of the outer planets. *Journal of Geophysical Research* 98: 18659–18679.
- Cowling TG (1934) The magnetic field of sunspots. *Monthly Notices of the Royal Astronomical Society* 34: 39–48.
- Crary FJ and Begenal F (1998) Remanent ferromagnetism and the interior structure of Ganymede. *Journal of Geophysical Research* 103: 25757–25773.
- Dolginov Sh Sh (1977) Planetary magnetism: A survey. *Geomagnetism and Aeronomy* 17: 391–406.
- Fearn DR (1979) Thermal and magnetic instabilities in a rapidly rotating fluid sphere. *Geophysical and Astrophysical Fluid Dynamics* 14: 103–126.
- Giampieri G and Balogh A (2002) Mercury's thermoelectric dynamo model revisited. *Planetary and Space Science* 50: 757–762.
- Gilbert N (1600) *De Magnete*, Gilbert Club Revised English Translation. London: Chiswick Press.
- Glatzmaier GA and Roberts PH (1995) A three-dimensional self-consistent computer simulation of a geomagnetic field reversal. *Nature* 377: 203–209.
- Glatzmaier GA and Roberts PH (1996) An anelastic evolutionary geodynamo simulation driven by compositional and thermal convection. *Physica D* 97: 81–94.
- Grote E and Busse FH (2001) Dynamics of convection and dynamos in rotating spherical fluid shells. *Fluid Dynamics Research* 28: 349–368.
- Grote E, Busse FH, and Tilgner A (2000) Regular and chaotic spherical dynamos. *Physics of the Earth and Planetary Interiors* 117: 259–272.
- Herzenberg A (1958) Geomagnetic dynamos. *Philosophical Transactions of the Royal Society of London A* 250: 543–585.
- Hide R (1974) Jupiter and Saturn. *Proceedings of the Royal Society of London A* 336: 63–84.
- Holme R and Bloxham J (1996) The magnetic fields of Uranus and Neptune: Methods and models. *Journal of Geophysical Research* 101: 2177–2200.
- Jacobs JA (1979) Planetary magnetic fields. *Geophysical Research Letters* 6: 213–214.
- Jones CA (2000) Convection-driven geodynamo models. *Philosophical Transactions of the Royal Society of London A* 358: 873–897.
- Jones CA (2003) Dynamos in planets. In: Thompson MJ and Christensen-Dalsgard J (eds.) *Stellar Astrophysical Fluid Dynamics*, pp. 159–176. Cambridge: Cambridge University Press.
- Jones CA, Soward AM, and Mussa AI (2000) The onset of thermal convection in a rapidly rotating sphere. *Journal of Fluid Mechanics* 405: 157–179.
- Kivelson MG, Khurana KK, Russell CT, et al. (1996) Discovery of Ganymede's magnetic field by the Galileo spacecraft. *Nature* 384: 537–541.

- Kutzner C and Christensen UR (2000) Effects of driving mechanisms in geodynamo models. *Geophysical Research Letters* 27: 29–32.
- Kutzner C and Christensen UR (2002) From stable dipolar towards reversing numerical dynamos. *Physics of the Earth and Planetary Interiors* 131: 29–45.
- Larmor J (1919) How could a rotating body such as the sun become a magnet? *British Association for the Advancement of Science Report* 159–160.
- Lee KKM, Benedetti LR, Jeanloz R, et al. (2006) Laser-driven shock experiments on precompressed water: Implications for ‘icy’ giant planets. *Journal of Chemical Physics* 125: 014701.
- Liu J, Goldreich PM, and Stevenson DJ (2007) Ohmic dissipation constraints on deep-seated zonal winds in Jupiter and Saturn. *Icarus* (in press).
- Love JJ (2000) Dynamo action and the nearly axisymmetric magnetic field of Saturn. *Geophysical Research Letters* 27: 2888–2892.
- Malkus WVR (1968) Precession of the Earth as the cause of geomagnetism. *Science* 160: 259–264.
- Malkus WVR (1994) Energy sources for planetary dynamos. In: Proctor MRE and Gilbert AD (eds.) *Lectures on Planetary and Solar Dynamos*, pp. 161–179. Cambridge: Cambridge University Press.
- Margot J, Peale S, Jurgens RF, Slade MA, and Holin IV (2004) Earth-based measurements of Mercury’s forced librations in longitude. *EOS Transactions of the American Geophysical Union* 85: GP33-03.
- Nellis WJ, Hamilton DC, Holmes NC, et al. (1988) The nature of the interior of Uranus based on studies of planetary ices at high dynamic pressure. *Science* 240: 779–781.
- Nellis WJ, Weir ST, and Mitchell AC (1996) Metallization and electrical conductivity of hydrogen in Jupiter. *Science* 273: 936–938.
- Runcorn S (1975) An ancient lunar magnetic field. *Nature* 253: 701–703.
- Ruzmaikin AA and Starchenko SV (1991) On the origin of Uranus and Neptune magnetic fields. *Icarus* 93: 82–87.
- Sarson GR, Jones CA, Zhang K, and Schubert G (1997) Magnetoconvection dynamos and the magnetic field of Io and Ganymede. *Science* 276: 1106–1108.
- Showman AP, Stevenson DJ, and Malhotra R (1997) Coupled orbital and thermal evolution of Ganymede. *Icarus* 129: 367–383.
- Simitev R and Busse FH (2005) Prandtl-number dependence of convection-driven dynamos in rotating spherical fluid shells. *Journal of Fluid Mechanics* 532: 365–388.
- Somerville RCJ and Lipps FB (1973) A numerical study in three space dimensions of Bénard convection in a rotating fluid. *Journal of Atmospheric Sciences* 30: 590–596.
- Soward AM (1974) A convection driven dynamo. Part I: The weak field case. *Philosophical Transactions of the Royal Society of London A* 275: 611–645.
- Stanley S and Bloxham J (2006) Numerical dynamo models of Uranus’ and Neptune’s magnetic fields. *Icarus* 184: 556–572.
- Stanley S, Bloxham J, Hutchison WE, and Zuber MT (2005) Thin shell dynamo models consistent with Mercury’s weak observed magnetic field. *Earth and Planetary Science Letters* 234: 27–38.
- Stegman DR, Jellinek AM, Zatman SA, Baumgardner JR, and Richards MA (2003) An early lunar core dynamo driven by thermochemical mantle convection. *Nature* 421: 143–146.
- Stevenson DJ (1979) Turbulent thermal convection in the presence of rotation and a magnetic field. *Geophysical and Astrophysical Fluid Dynamics* 12: 139–169.
- Stevenson DJ (1982) Reducing the non-axisymmetry of a planetary dynamo and an application to Saturn. *Geophysical and Astrophysical Fluid Dynamics* 21: 113–127.
- Stevenson DJ (1983) Planetary magnetic fields. *Reports on Progress in Physics* 46: 555–620.
- Stevenson DJ (1984) The energy flux number and three types of planetary dynamo. *Astronomische Nachrichten* 305: 257–264.
- Stevenson DJ (1987) Mercury’s magnetic field: A thermoelectric dynamo? *Earth and Planetary Science Letters* 82: 114–120.
- Stevenson DJ (2001) Mars’ core and magnetism. *Nature* 412: 214–219.
- Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11.
- Tilgner A (1999) Spectral methods for the simulation of incompressible flows in spherical shells. *International Journal of Numerical Methods in Fluids* 30: 713–724.
- Tilgner A (2005) Precession driven dynamos. *Physics of Fluids* 17: 034104-1–6.
- Tilgner A and Busse FH (1997) Finite amplitude convection in rotating spherical fluid shells. *Journal of Fluid Mechanics* 332: 359–376.
- Vanyo JP (1984) Earth core motions: Experiments with spheroids. *Geophysical Journal of the Royal Astronomical Society* 77: 173–183.
- Zhang K (1994) On coupling between the Poincaré equation and the heat equation. *Journal of Fluid Mechanics* 268: 211–229.
- Zhang K (1995) On coupling between the Poincaré equation and the heat equation: No-slip boundary condition. *Journal of Fluid Mechanics* 284: 239–256.

10.09 Dynamics and Thermal History of the Terrestrial Planets, the Moon, and Io

D. Breuer, DLR Institute of Planetary Research, Berlin, Germany

W. B. Moore, University of California, Los Angeles, CA, USA

© 2007 Elsevier B.V. All rights reserved.

10.09.1	Introduction	299
10.09.2	Physical and Chemical Properties of Planets and Planetary Materials Bearing on Mantle Dynamics and Thermal Evolution Models	300
10.09.2.1	Interior Structure and Phase Transitions	300
10.09.2.2	Heat Sources	301
10.09.2.2.1	Accretional heating	301
10.09.2.2.2	Core formation	302
10.09.2.2.3	Radioactive heat sources	303
10.09.2.2.4	Tidal dissipation	303
10.09.2.3	Rheology	303
10.09.2.4	Thermal Expansivity and Conductivity	304
10.09.3	Planform of Convection	305
10.09.4	Thermal Evolution Models Using Parametrized Convection	306
10.09.4.1	Parametrizations Based on Constant-Viscosity Laws	307
10.09.4.2	Stagnant Lid Parametrization Based on Temperature-Dependent Viscosity Laws	308
10.09.4.3	Plate Tectonics, Lithosphere Delamination, and Stagnant Lid Convection	310
10.09.5	Thermal Evolution, Volcanic History, and Magnetic Field History of Terrestrial Planets	312
10.09.5.1	Crust Formation	313
10.09.5.1.1	Primary and secondary crust formation	313
10.09.5.1.2	Plume volcanism and partial melt zone	314
10.09.5.1.3	Thermal properties of the crust	315
10.09.5.1.4	Observations	315
10.09.5.1.5	Incorporation of crustal growth in thermochemical evolution models	316
10.09.5.2	Magnetic Field Generation	316
10.09.5.2.1	Core convection and necessary conditions for dynamos	317
10.09.6	Comparison of the Terrestrial Planets and the Moon	318
10.09.6.1	Mercury	318
10.09.6.2	Venus	322
10.09.6.3	Moon	325
10.09.6.4	Mars	329
10.09.7	Io	335
10.09.7.1	Tidal Heating and the Evolution	335
10.09.7.2	Thermal and Orbital Evolution	337
10.09.8	Summary	339
References		341

10.09.1 Introduction

Unraveling a planet's evolution in order to understand the processes and features observed on the terrestrial planets is a difficult task that requires the integration of evidence from various fields into

consistent models. Geology, geophysics, mineralogy, and cosmochemistry, as well as chemistry, physics, and biology, all influence the thermal and chemical evolution of terrestrial bodies. Our knowledge of the other terrestrial bodies stems mostly from remote-sensing satellites, but in some cases also from *in situ*

chemical measurements, and even from surface samples (in the case of the Moon) and meteorites (in the case of Mars). These samples have been of immense value for Moon and Mars science as they provide information not only about the composition but also about processes like the formation of the core and the crust of those bodies. A fundamental process in planetary evolution is the thermal evolution of its interior, which depends on the internal heat sources and, to a large extent, on the dynamics of the planet's mantle through which heat is transported by thermal convection. It has been understood since the beginning of the twentieth century that thermal convection is capable of driving mantle convection. If the mantle is heated from within (or from below) and is cooled from above, it becomes gravitationally unstable and thermal convection can occur as colder rock descends into the mantle and hotter rock ascends toward the surface. This circulation of material transports heat toward the planet's surface and tends to cool the interior, while heat produced within, for example, by the decay of radioactive elements tends to warm it. The motions driven by convective heat transport result in surface stresses and deformation, producing the geologic features observed on the terrestrial bodies today. An understanding of the thermal evolution is critical for interpreting these surface features in terms of interior processes including the dynamics of the mantle and the core.

The Earth with its plate tectonics is unique among the terrestrial planets. Our planet's segmented lithosphere forms along ridges under the oceans, moves laterally outward at geologically rapid rates, and is destroyed (subducted) at linear boundaries. This recycling process transfers heat out of the interior very effectively. One consequence of this relatively rapid process is that the Earth has a very young oceanic crust of only 60 Ma. In addition to the oceanic crust, a less-dense continental crust exists that usually does not recycle and has an average age of about 2000 Ma. Comparing the Earth with the other terrestrial planets, the most striking difference at first glance is the state of the surface (including the crust) and the way heat is transported. The surfaces of other planets are not segmented but consist of single plates, the so called stagnant lids, beneath which the mantle convects. Heat flow through the surface is mainly transported by conduction, with some minor contribution by volcanic heat transport through this stagnant lid. The difference in the heat transport

mechanism for the planets is also reflected in their thermal evolution.

In this chapter, we describe the thermal evolution and the mantle dynamics as well as related processes like crustal formation and the magnetic field evolution of the terrestrial bodies other than Earth, that is, Mercury, Venus, Mars, and the Moon. The Galilean satellite Io is included in this chapter as this body consists, like the terrestrial planets and the Moon, of an iron-rich core and a silicate outer shell. The close orbit about Jupiter and the resonant orbital interaction among Io, Europa, and Ganymede make Io, however, special among the other terrestrial bodies. In its interior, tidal deformation plays an important role and provides an extremely large internal heat source. The consequence is a very different thermal evolution for Io than for the other bodies; therefore, we discuss Io in a separate section of this chapter.

The chapter is organized as follows: First, we describe the relevant physical and chemical properties of planets and planetary materials bearing on mantle dynamics and thermal evolution models. In the next section the concept of mantle convection is introduced and the relationships between convective parameters and the planform of the flow are discussed. A description of parametrized convection models follows with a comparison of three main heat transport mechanisms: plate tectonics, stagnant lid convection, and lithosphere delamination. In the next section crustal formation and magnetic field generation of terrestrial planets is described since both strongly depend on the thermal evolution of a planet and can be used as constraints for the models. The chapter concludes with separate reviews of the thermochemical evolution and mantle dynamics of Mercury, Venus, Moon, Mars and Io, and a summary section. For a discussion of the thermal evolution of icy satellites, see Chapter 10.15.

10.09.2 Physical and Chemical Properties of Planets and Planetary Materials Bearing on Mantle Dynamics and Thermal Evolution Models

10.09.2.1 Interior Structure and Phase Transitions

The interior dynamics and thermal evolution of the terrestrial planets depend strongly on internal structure and composition. One important piece of information is the size and density of the major reservoirs, that is, core, mantle, and crust. Lacking

seismic data for most bodies, the basic structure can be estimated from the mean density and the polar moment of inertia (MoI) factor of the terrestrial bodies. The latter parameter can be obtained from measurements of the low-order gravity field (J_2 and C_{22}), the figure and/or the precession rate of the planet. The use of shape or gravity data to infer the MoI requires a planet to be in hydrostatic equilibrium. The determination of the interior structure does not provide a unique solution; there is, for example, a tradeoff between core size and core density. However, with additional constraints on composition, it is possible to derive estimates of the structure (see Chapter 10.02). For mantle dynamics, the size of the metallic core is of particular interest. A large core (and hence a thin mantle), suggests a more efficient cooling of the mantle layer and one finds in general a convection pattern that is smaller in scale compared to a planet with a thick mantle.

Purely thermal convection can be strongly influenced by chemical layering and phase transitions in the mantle. In the case of chemical layering, density variations that drive convection are not only caused by temperature variations but also by differences in mantle composition. Chemical layering can block or retard mantle flow if the compositional density increases with depth. In this stable configuration, thermal convection is only possible if density changes due to temperature variations can overcome the compositional density differences. Chemical layering can enhance mantle flow if the compositional density decreases with depth. Reshuffling of this unstable configuration results in a stable density configuration unless efficient mixing prevents the establishment of distinct layers. It is also possible that thermal convection takes place in separate and chemically distinct layers. The heat transport in a mantle with layered thermal convection is less efficient than in a mantle with whole-mantle convection. Chemical stratification resulting from early differentiation of the mantle is one way to explain the presence of separate and chemically distinct reservoirs that might be needed to explain geochemical observations. At present, we do not know whether distinct chemical reservoirs exist in the terrestrial planets and the Moon.

Phase transitions in planetary mantles also play an important role in the dynamics and thermal evolution of the planet. At a certain pressure and temperature mantle materials may undergo isochemical phase changes, associated with a change in density. The variation of the phase-transition temperature with pressure is related to the volume and

entropy changes through the Clausius–Clapeyron equation. The main phase transitions in an olivine-rich terrestrial mantle are the exothermic olivine–spinel and the endothermic spinel–perovskite transitions. Depending on the pressure gradient within a planet, these transitions are located at different depths. For instance, in the Earth's mantle the olivine-to-spinel transitions are located at a depth of 400–520 km whereas in the smaller Mars the depth of this transition zone is roughly between 1000 and 1500 km. Recent mineral physics experiments and *ab initio* calculations (Murakami *et al.*, 2004; Tsuchiya *et al.*, 2004; Oganov and Ono, 2004) have indicated that also perovskite undergoes an exothermic phase change to a post-perovskite structure just above the core–mantle boundary (CMB) of the Earth. For Mercury and the Moon the pressure gradient is too low, and olivine–spinel phase-transition pressures are not reached in their mantles. Details about the interior structure of the terrestrial planets and the Moon are given in Chapter 10.02.

10.09.2.2 Heat Sources

The planet's internal heat sources provide the energy for its dynamics, supplying it with the driving energy for mantle convection, and for melting the mantle which may lead to volcanism. The internal heat energy of the terrestrial planets was greater in the early stages of their histories than it is today, having accumulated rapidly by heat conversion associated with three separate processes, all of which were most intense during the first few million years of the planet's history: (1) accretion of the planet by impacts, (2) core formation and the associated release of gravitational potential energy, and (3) the radioactive decay of unstable isotopes. Another important heat source in a terrestrial body can be tidal dissipation, as will be seen in the case of Io, below. This heat source is not necessarily higher early in the evolution; rather, it depends on the body's orbital and thermal evolution.

10.09.2.2.1 Accretional heating

Accretion of material during the formation of planets is likely one of the largest sources of heat. The accretional heat is the energy accumulated during the burial of heat by impacts as the planet grows through the accretion of planetesimals. During accretion, a temperature profile is generated for which the temperature increases from the center toward the surface (Figure 1). A possible postaccretional

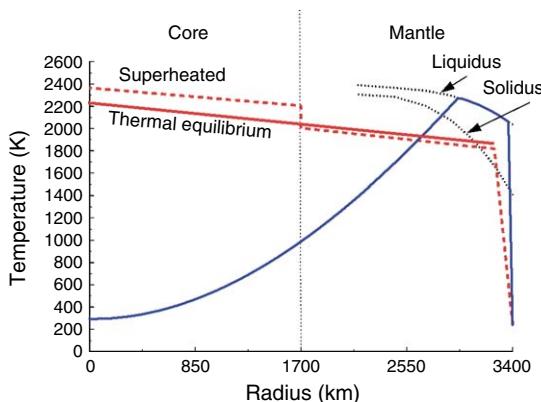


Figure 1 Accretional temperature profile for a planet with the size of Mars (blue solid line) calculated from eqn [1] with a retention factor h of 0.2. The accretional temperature exceeds the solidus and the liquidus close to the surface, thus forming a 600-km-deep magma ocean. After core formation two possible temperature profiles can develop if (1) the core formed in thermal equilibrium with the mantle (solid red line) and (2) the core is superheated with respect to the mantle as the consequence of rapid core formation (dashed red line). The temperature profiles account for the potential energy release due to core formation (eqn [4]). This energy is distributed homogeneously assuming an adiabatic interior.

temperature profile for homogeneous accretion (e.g., Schubert *et al.*, 1986) is given by

$$T_a(r) = b \frac{GM(r)}{C_p r} \left(1 + \frac{ru^2}{2GM(r)} \right) + T_e + \Delta T_{a,ad} \quad [1]$$

where b is the retention factor describing the contribution of heat retained in the accumulated body, G is the gravitational constant, $M(r)$ is the mass of the planet internal to radius r , C_p is the specific heat, $u^2/2$ is the approach kinetic energy per unit mass, and T_e is the temperature of the surrounding nebula. The postaccretional temperature profile of Schubert *et al.* (1986) is corrected to account for the temperature rise, $\Delta T_{a,ad}$, caused by adiabatic compression as the internal pressure progressively increases during accumulation.

This accretion temperature distribution is stable against thermal convection and shows an inverse profile, which decreases with depth (Figure 1) when compared to a typical temperature profile for planetary interiors, which increases with depth. Depending on the amount of stored heat (part of the energy is radiated back to space) and, in particular, on the size of the body, the accretion profile can rise above the melting temperature of iron and possibly of silicate close to the surface; a magma ocean may form.

10.09.2.2.2 Core formation

Planets are assumed to have initially accreted as a mixture of silicate and metal particles. It is, however, likely that at the late stage of accretion the planetesimals are already differentiated (e.g., Kleine *et al.*, 2004; Baker *et al.*, 2005). The separation of the metal particles, that is, the formation of an iron-rich core, most likely was contemporaneous with or shortly after a planet's accretion. Recent studies on short-lived radio-nuclides (e.g., ^{182}Hf) suggest that the separation between silicates and iron occurred on the order of $1-5 \times 10^7$ years after accretion (e.g., Kleine *et al.*, 2002). Estimates of the timescale of the core formation process suggest that for rapid separation both the silicate and the iron need to be fluid at least in the upper part of the planet (e.g., Stevenson, 1990). In such a scenario, iron accumulates at the base of a magma ocean and sinks further to the center of the planet by Rayleigh–Taylor instability. Alternatively, it has been assumed that no global magma ocean formed during accretion; rather, silicate–metal separation takes place without global-scale melting but in magma ponds generated by impacts (Tonks and Melosh, 1992). Numerical studies of this process, however, suggest that the time to form a core would be about 0.5–1 Ga (Senshu *et al.*, 2002).

The gravitational energy released by core formation is converted into thermal energy, which strongly heats the interior. During core formation, when iron sinks to the centre, the temperature profile is inverted to decrease from the center toward the surface. The total energy released by the differentiation of a homogeneous planet into an iron-rich core and a silicate mantle can be estimated from the difference between the potential energy stored in a homogeneous planet after accretion, U_H , and the potential energy of the differentiated, two-layer planet, U_D (Schubert *et al.*, 1986). Thus, the energy dissipated by differentiation is

$$U_{\text{Diff}} = U_H - U_D \quad [2]$$

with the potential energy stored in a homogeneous satellite after accretion

$$U_H = - \frac{16}{15} \pi^2 G R_p^5 \rho^2 \quad [3]$$

and with the potential energy of a two-layer planet (with constant density ρ_m in the mantle and ρ_c in the core) given by

$$U_D = -\frac{16}{15}\pi^2 GR^5 \left[\rho_m^2 + \frac{5}{2}\rho_m(\rho - \rho_m) + \left(\frac{3}{2}\rho_m - \rho_c \right) (\rho_m - \rho_c) \left(\frac{\rho - \rho_m}{\rho_c - \rho_m} \right)^{5/3} \right] \quad [4]$$

If this energy is homogeneously distributed in the planetary interior, that is, if there is thermal equilibrium between the mantle and the core, estimates for the mean temperature increase are about 1000 K for the Earth and about 300 K for Mars (Solomon, 1979). It is, however, likely that due to the assumed rapid core formation process the core was superheated with respect to the mantle (Figure 1). The excess temperature might be the consequence of adiabatic heating of the sinking iron since the core alloy has a thermal expansion coefficient which is 2–3 times larger than that of mantle silicates and might be larger for larger bodies. Furthermore, the one order of magnitude higher thermal conductivity of iron in comparison to silicate allows more efficient heating of the sinking iron blobs by viscous dissipation than of the silicate mantle through which the iron sinks.

10.09.2.2.3 Radioactive heat sources

The two heat sources just discussed are only relevant at the early time when accretion and core formation occurs. Since one expects these processes to be very rapid, they only provide a pulse of heat early in the history and thus they are not responsible for remelting or tectonics after the first few hundred million years. Instead, long-lived radioactive elements with half-lives of billions of years, that is, ^{238}U , ^{235}U , ^{232}Th , ^{40}K , become the dominant heat sources. It is this heat that can produce volcanoes and power internal convection and tectonics on planets. At present, roughly half of the Earth's heat output is thought to be due to radioactive cooling, while the remainder may be due to secular cooling (loss of sensible heat). For the other terrestrial planets this ratio might be similar. The distribution of the radioactive heat source can change with time as a consequence of chemical differentiation (crust formation). The radiogenic elements are incompatible in the solid mantle undergoing melting and are thus enriched in the melt. This melt then rises to the surface and forms new crust. If the radioactively enriched crust is stably stratified at the surface, the mantle heat sources are reduced.

Short-lived isotopes, in particular ^{26}Al with a half-life of 0.74 Ma, might also have played an important role in heating rocky bodies when the solar system

was first forming (e.g., Gosh and McSween, 1998; Merk *et al.*, 2002). The heat was high at an early stage but disappeared quickly. Possibly, this strong energy peak was responsible for the differentiation of some of the asteroids and of planetesimals that formed the planets.

10.09.2.2.4 Tidal dissipation

An important heat source in terrestrial planets can be tidal energy caused by gravitational interactions between the sun and the planets and/or the planets and their satellites. For a satellite that moves on an elliptical orbit, the distance and therefore the gravitational attraction varies periodically. If the interior reacts to this periodic deformation viscously and not elastically, tidal energy can be dissipated in the form of heat. The heating due to tidal dissipation can actually be higher than the heat released by the decay of radioactive elements. The tidal dissipation rate, \dot{E} , depends on the time variation of the tides, the interior structure, and the rheology (and hence, temperature) of the dissipating body. The orbital evolution is therefore coupled to the thermal evolution by the tidal dissipation rate. In the case of a small inclination of the equatorial plane to the orbital plane and a small eccentricity, the tidal dissipation rate can be estimated using the relation by Segatz *et al.* (1988):

$$\dot{E} \sim \text{Im}(k_2)_{\text{Sat}} \frac{(nR_{\text{Sat}})^5}{G} e^2 \quad [5]$$

with R_{Sat} the radius of the satellite, n the mean motion, e the eccentricity, $\text{Im}(k_2)_{\text{Sat}}$ the imaginary part of the tidal Love number k_2 (measuring the phase lag of the tide), and G the gravitational constant.

10.09.2.3 Rheology

Rheology is a primary factor affecting convection in terrestrial planets and a rheological prerequisite for convection is that the material be a fluid, that is, able to flow under small differential stresses. The time factor is of utmost importance. At high frequency the mantle is approximately elastic, but at very low frequency (periods larger than 1000 years) the mantle (at sufficient high temperatures and pressure) behaves like a viscous fluid. Unfortunately, such conditions cannot be reproduced in the laboratory and thus the true rheology of planetary mantles is poorly known. What is known is that the viscosity of mantle rocks under the wide variety of conditions in planetary mantles is dependent on temperature, pressure, stress, grain size, and composition.

The exponential dependence of the rheology on the inverse absolute temperature is particularly important in understanding the role of mantle convection in transporting heat. The temperature dependence of the rheology acts as a thermostat to regulate the mantle temperature. Any tendency of the mean temperature to increase is offset by an associated reduction in mantle viscosity, an increase in convective vigor, and a more efficient outward transport of heat. Similarly, a decrease in mantle temperature tends to increase mantle viscosity, reduce convective flow velocities, and decrease the rate of heat transfer. As a result of the sensitive feedback between mantle temperature and rheology, relatively small changes in temperature can produce large changes in heat flux, and the temperature is consequently buffered at nearly constant temperature (Tozer, 1965).

The viscosity of the mantle can be written in a simple Arrhenius form as

$$\eta = \eta_0 \exp\left(\frac{A + pV}{RT_m}\right) \quad [6]$$

In eqn [6], η_0 contains the stress, composition, and grain size-sensitive terms, T_m is the mean temperature of the convecting fluid, R is the universal gas constant, A is the activation energy for creep (e.g., Weertman and Weertman, 1975), p is the pressure, and V is the activation volume.

Though planetary mantles are composed of several minerals, olivine is thought to dominate the rheology of silicate mantle rock (e.g., Mackwell, 1991). Its viscosity is about 10^{21} Pa s at a reference temperature of 1600 K (e.g., Karato and Wu, 1993). The activation energy for dry olivine is about 540 kJ mol⁻¹ at pressures of 12 GPa (Karato and Wu, 1993) and the activation volume is about 15 cm³ mol⁻¹ (Karato and Rubie, 1997). The viscosity is also known to strongly depend on the presence of volatiles. Depending on the water content, the viscosity of olivine can vary between 10^{19} and 10^{21} Pa s at a reference temperature of 1600 K for either a weak and wet rheology or a stiff and dry rheology, respectively. The activation energy for wet olivine is 240 kJ mol⁻¹ (Karato and Wu, 1993).

10.09.2.4 Thermal Expansivity and Conductivity

Two important parameters for mantle dynamics and thermal evolution of terrestrial planets are the thermal expansivity and the thermal conductivity. The

former parameter describes the volume increase per unit temperature increase, that is, the decrease of density with increase of temperature. The thermal conductivity k is a bulk property of material that indicates its ability to conduct heat.

In terrestrial mantles, neither of the parameters is constant although thermal evolution models usually assume a constant value. Thermal expansivity decreases with pressure. Chopelas and Bohler (1989) discovered experimentally that for olivine and magnesium oxides, thermal expansivity decreases with the fifth or sixth power of density up to pressures of about 33 GPa. Theoretical estimates of Anderson (1987) suggest a slower decrease with the third power of density. Since the thermal buoyancy force driving convection and the adiabatic cooling are both proportional to the thermal expansivity, one can expect this variation to have an important influence on the convection pattern in planetary mantles where the mantle thickness is not much smaller than the density scale height.

Thermal conductivity depends on various factors like temperature, pressure, composition, and texture of the material. In the mantle, the phonon contribution, k_{lat} , and the radiative contribution k_{rad} contribute to the thermal conductivity

$$k_m(T, p) = k_{\text{lat}}(T, p) + k_{\text{rad}}(T) \quad [7]$$

The phonon contribution k_{lat} decreases with increasing temperature T and increases with increasing pressure p , whereas the radiative contribution k_{rad} increases with increasing temperature independent of pressure; for relevant temperatures in a terrestrial mantle, the phonon conductivity is much higher than the radiative contribution. As a consequence, a decrease of the thermal conductivity through the upper part of the mantle, that is, the stagnant lid and the upper thermal boundary layer, and an increase through the mantle beneath is expected. Hofmeister (1999) has developed a model for the temperature and pressure dependence of mantle thermal conductivity under terrestrial conditions:

$$k_{\text{lat}}(T, p) = k_{298} \left(\frac{298}{T}\right)^{\alpha} \exp\left[-\left(4\gamma + \frac{1}{3}\right)\alpha(T - 298)\right] \times \left(1 + \frac{K_T p}{K_T}\right) \quad [8]$$

$$k_{\text{rad}}(T) = 0.01753 - 0.00010365 T + 2.2451 T^2 / 10^7 - 3.407 T^3 / 10^{11} \quad [9]$$

where k_{298} is the surface thermal conductivity, α is the thermal expansivity and is taken to be constant, γ

is the Grüneisen parameter, K_T is the bulk modulus, and $K'_T = dK_T/dP$. The value of α depends on the chemical bonding, with $\alpha=0.3$ characteristic of silicates and $\alpha=0.9$ characteristic of oxides.

In parametrized convection models, it is often assumed that the thermal conductivity in the terrestrial mantle has a constant value of $4 \text{ W m}^{-1} \text{ K}^{-1}$ (e.g., Stevenson *et al.*, 1983; Spohn, 1991; Schubert *et al.*, 1992; Breuer and Spohn, 2003). Applying eqns [8] and [9] to terrestrial planets, however, shows that the thermal conductivity is actually lower for a large part of the mantle if not the entire mantle. Due to the strong temperature dependence, the mantle thermal conductivity is actually lower in the early evolution. For instance in Mars, values may range between 2.5 and $3.5 \text{ W m}^{-1} \text{ K}^{-1}$ at the bottom of the upper thermal boundary layer (Schumacher and Breuer, 2006).

10.09.3 Planform of Convection

Convection is driven by density variations which arise from thermal as well as chemical variations. Here we will be primarily concerned with thermal convection. The Rayleigh number describes the degree of instability and the convective vigor of a convecting system. With increasing convective vigor thermal boundary layers get thinner and the convection planform smaller (Turcotte and Schubert 1982); thereby, the horizontal scale of instabilities is reflective of the thickness of the layer itself. In general the scale of convection depends on the geometry of the system, the heating mode, and the mechanical and thermal boundary conditions. For planetary mantles shear-stress-free conditions with an isothermal top boundary are appropriate, while the heating mode can vary from purely basally heated to purely internally heated whereas any mixed heating mode in between is possible.

Figure 2 shows the critical Rayleigh number (the Rayleigh number at which convection sets in) for bottom heated (a) and purely internally heated (b) spherical shells, as a function of the aspect ratio of the shell ($\eta = \text{inner radius}/\text{outer radius}$) for different axisymmetric modes of convection, denoted by spherical harmonic degree l (Zebib *et al.*, 1983). As the shell becomes thinner (η increasing), the harmonic degree of the most unstable mode increases, and the critical Rayleigh number for instability decreases, approaching the plane layer values as η approaches 1. Thus, Mercury's mantle with $\eta=0.75$ has the lowest critical

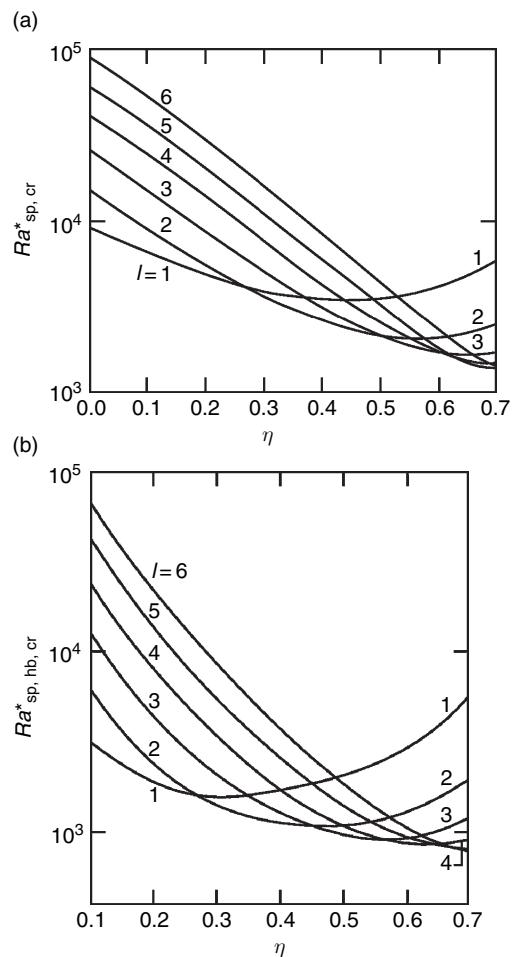


Figure 2 Critical Rayleigh number for bottom heated (a) and purely internally heated (b) spherical shells, as a function of the aspect ratio η of the shell. The curves for different modes are labeled by spherical harmonic degree l . Modified from Zebib A, Schubert G, Dein JL, and Paliwal RC (1983) Character and stability of axisymmetric thermal convection in spheres and spherical shells. *Geophysical and Astrophysical Fluid Dynamics* 23: 1–42 (<http://www.informaworld.com>).

Rayleigh number of all the terrestrial planets. Nonetheless, Mercury's small size leads to a low intrinsic Rayleigh number for the shell and it is not clear if convection occurs at all (e.g., Hauck *et al.*, 2004).

For small core sizes (low η), the most unstable modes have very low l , and for η below about 0.25, the degree-1 mode is the most unstable. This mode has a single, cylindrical upwelling; and a broader, weaker downwelling located antipodally. It was thought for some time that this convection mode described the Martian interior, with the huge volcanic upland of Tharsis located above the upwelling.

A recent measurement of the MoI of Mars (Folkner *et al.*, 1997), however, found that the aspect ratio of the Martian mantle is much larger than 0.25, and other solutions for a degree-1 mode were found involving phase transitions (see below). Higher-degree modes have more upwellings and downwellings, with either taking the form of sheets or columns depending on the polarity of the solution. Sheet-like downwellings seem to be the preferred geometry for planetary mantles, but upwellings may also organize along linear structures (Schubert *et al.*, 2001), perhaps explaining chains of volcanic highlands such as those seen on Venus (Ratcliff *et al.*, 1997) or the line of large volcanoes in the Tharsis region of Mars.

For a given shell aspect ratio, increasing the Rayleigh number above the critical value causes additional modes to become unstable, some larger scale (smaller l) and some smaller scale. As the Rayleigh number continues to increase, more and more high-degree modes become unstable, and the scale of these modes becomes small compared with the depth of the layer. At this point, the cellular nature of the convection breaks down, and the flow is better described by the transient formation of small-scale instabilities in the boundary layers. This happens for Rayleigh numbers of 10^7 – 10^8 , which is close to the values expected for the mantles of Venus and Earth. It has been suggested that the volcanic coronas unique to Venus are the surface expression of such instabilities (Stofan *et al.*, 1991).

Another factor which influences the planform of convection is the presence of mantle phase changes. In Venus and Earth, at least two significant solid-state phase changes associated with the transformation of olivine to spinel and spinel to perovskite structure are present. The olivine–spinel transition is exothermic and thus can act to enhance convective motions, while the spinel–perovskite phase change is endothermic and thus can act to retard convection (Schubert *et al.*, 1970). The influence of phase changes on convective planform has two aspects: first, if the endothermic phase change can act as a significant barrier to convection, the layer may be split into two independently (or nearly so) convecting layers, each with a higher aspect ratio and a lower Ra than the whole mantle. Second, the effectiveness of the endothermic phase change in resisting convective motions is wavelength dependent, becoming less effective as the wavelength increases (Bercovici *et al.*, 1993). Thus this transition can act as a filter, preventing short-wavelength instabilities from passing but allowing larger-scale motions (Tackley *et al.*, 1993,

1994; Tackley, 1996). For Venus, this perhaps explains the large-scale structure of the highland regions superimposed on the small-scale features that generate coronas. In Mars, the olivine–spinel transition falls near the middle of the mantle, but the spinel–perovskite transition, if it is reached at all, is located very near the core. Convection calculations with this geometry show that the number of upwellings originating from the CMB is reduced by the presence of the endothermic phase transition and that single-plume planforms are possible (Harder and Christensen, 1996; Harder, 1998).

Convective planform is also influenced by rheological stratification. Although we know very little of the interior structure of the other terrestrial planets, the Earth shows a strong rheological contrast between upper and lower mantle, with estimates ranging from a factor of 30–1000 in the viscosity difference between the upper and lower mantle (e.g., Peltier and Jiang, 1996; Kaufmann and Lambeck, 2000; Cadek and Fleitout, 2003). Additionally, the viscosity of mantle materials shows a continuous increase with pressure. Finally, the Earth exhibits a global or nearly global low-viscosity zone (asthenosphere) beneath the lithosphere. The rheological stratification of the Earth affects the length scales of convection in two ways: first, the high-viscosity lower mantle has an effective Rayleigh number that is lower than the mantle as a whole and thus the scale of instabilities will increase relative to the upper mantle. Second, the low-viscosity zone acts to decrease the dissipation associated with shear across the upper thermal boundary layer, allowing longer cells to form (Lenardic *et al.*, 2006; Busse *et al.*, 2006). It is not clear if either of these effects is present in the other terrestrial planets. The low-viscosity zone, in particular, is thought to be related to the recycling of water into the mantle by subduction, a process that is at present not relevant to other bodies.

10.09.4 Thermal Evolution Models Using Parametrized Convection

Thermal evolution calculations with 2-D or 3-D mantle convection codes are very time consuming on present-day computers. In these calculations, a set of differential equations has to be solved for the conservation of mass, momentum, and energy including an appropriate constitutive equation. Because of the inherent complexity in these models it is often

desirable to take an empirical approach and parametrize the convective heat transfer rate as a function of known quantities. Such parametrizations can be derived using simple theories, which result in scaling laws that describe the heat transport in the interior as a function of the convective parameters.

Our improving understanding of the heat transport mechanisms on terrestrial planets over the last two decades has led to repeated changes in the preferred scaling law used to model the thermal evolution of one-plate planets. Initially, the scaling law for a fluid with constant viscosity was used for one-plate planets. In the meantime, it has been recognized that this scaling law models the heat transport in a planet where convection comprises the whole mantle including the outer layers. In fact, such a model describes the heat transport in a planet with plate tectonics better than that in a one-plate planet (e.g., Schubert *et al.*, 2001). The plate tectonics regime is expected to cool the planet very efficiently because the comparatively cold outer layers become recycled into the interior of the mantle by convection (**Figure 3**).

Subsequent attempts to model the heat transport in one-plate planets included the effects of a growing lithosphere in the models with parametrizations

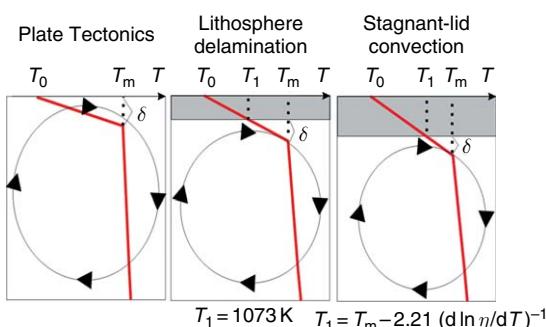


Figure 3 Sketch of the convection flow (circle with arrows) and the temperature profiles (red lines) in a mantle for three different heat transport mechanisms. In the plate tectonics regime, mantle convection comprises the entire mantle including outer layers. For this regime, the temperature difference across the thermal boundary layer (δ), which drives the convection, is given by $T_m - T_0$ (with T_m the interior temperature of the convecting mantle below the thermal boundary layer and T_0 the surface temperature). In the stagnant lid model and the lithosphere delamination model, mantle convection occurs below a lid (shaded area). The temperature difference across the thermal boundary layer is given by $T_m - T_\ell$ (with T_ℓ the temperature at the base of the lid). T_ℓ differs between these two models, being constant for the lithosphere delamination model and dependent on T_m and the mantle rheology for the stagnant lid model.

based on constant-viscosity scaling laws (e.g., Schubert *et al.*, 1990). The base of the nonconvecting lithosphere is represented by an isotherm characteristic of the transition from viscous deformation to rigid response to loads applied over geologic time-scales (e.g., Schubert *et al.*, 1992). Such a model represents the heat transport in a planet with a single plate on top of a convecting mantle. However, this model assumes that the lid coincides with the rheological lithosphere, that is, the region of the planets crust and mantle over which the flow properties of rock are insufficient to relieve stress. The part of the upper mantle that is weaker than the rheological lithosphere is assumed to be constantly recycled within the mantle. In comparison to the recently derived stagnant lid parametrization, these models, therefore, represent a mechanism of ‘lithosphere delamination’. The efficiency of the heat transport by lithosphere delamination is in-between that of plate tectonics and that of stagnant lid convection.

Recently, new scaling laws have been derived from convection models in layers of fluids with strongly temperature-dependent viscosities (e.g., Richter, 1978; Davaille and Jaupart, 1993; Solomatov, 1995; Moresi and Solomatov, 1995; Grasset and Parmentier, 1998; Reese *et al.*, 1999). These scaling laws have been suggested to represent the heat transport in a planet with a single plate on top of a convecting mantle. This stagnant lid convection predominantly cools the outer layers of a planet through stagnant lid growth while the deep interior is cooled rather inefficiently.

In the following, various approaches to model the heat transport that are either based on constant (Section 10.09.4.1) or temperature-dependent (Section 10.09.4.2) viscosity fluids are introduced.

10.09.4.1 Parametrizations Based on Constant-Viscosity Laws

From convection experiments and boundary layer theory for isoviscous fluids the classical power-law relationship between heat flow and convective parameters (e.g., Turcotte and Oxburgh, 1967; Roberts, 1979) has been derived:

$$Nu = \alpha Ra^\beta \quad [10]$$

This equation relates the dimensionless heat flux out of the convecting layer, expressed as the Nusselt number, Nu , to the strength of thermal convection as measured by the Rayleigh number, Ra .

The parameters α and β are constants, with α typically ranging between 0.195 and 0.339 and β between 1/4 and 1/3 depending on the geometry of the convecting layer and the boundary conditions (e.g., Jarvis, 1984; Schubert and Anderson, 1985; Zebib *et al.*, 1985).

The Nusselt number is a measure of the ability of convection to transfer heat in comparison with thermal conduction. It is calculated by dividing the average surface heat flux, q_s , by the theoretical heat flux, $k\Delta T_d/d$ where d is the thickness of entire fluid layer, k is the thermal conductivity, and ΔT_d is the temperature jump across the fluid layer:

$$Nu = q_s \frac{d}{k\Delta T_d} \quad [11]$$

For convection, the surface heat flow is given by

$$q_s = k \frac{\Delta T_\delta}{\delta} \quad [12]$$

where δ is the thickness of the upper thermal boundary layer, and ΔT_δ is the temperature difference across this boundary layer. With convection, Nu is larger than one. The Rayleigh number is defined as

$$Ra = \frac{\alpha \rho g \Delta T_d d^3}{\kappa \eta} \quad [13]$$

where α is the thermal expansivity, ρ is the density, κ is the thermal diffusivity, g is the acceleration due to gravity, and η is the viscosity of the fluid.

The power law in eqn [10] has been applied to model the thermal evolution of terrestrial planets, for example, of the Earth with a plate tectonic regime but also of one-plate planets such as Mercury, Venus, and Mars (e.g., Sharpe and Peltier, 1978; Schubert 1979; Schubert *et al.*, 1979a; Stevenson *et al.*, 1983). Equation[10] has been used although the viscosity η of the terrestrial mantles is thought to be strongly temperature dependent (eqn [6]).

The models for one-plate planets have later been modified to include the effects of a growing lithosphere by Schubert (1990) (see also Schubert *et al.* (1992) and Spohn (1991)) using an equation for the growth of a lithosphere, which was first proposed by Schubert *et al.* (1979b):

$$q_m - q_l = \rho_m C_m (T_m - T_l) \frac{dl}{dt} \quad [14]$$

where q_m and q_l denote the heat flow into the lithosphere from the convecting mantle and the heat flow through the lithosphere by conduction, respectively, C_m is the heat capacity, T_l is the temperature at the

base of the lithosphere, T_m is the temperature in the mantle at the base of the upper thermal boundary layer, l is the thickness of the lithosphere, and t is the time. The base of the lithosphere in these models is an isotherm assumed to be characteristic for the transition from viscous deformation to rigid response to loads applied over geologic timescales. This lithosphere can be identified with a rheological lithosphere. A representative value of T_l is around 1073 K.

10.09.4.2 Stagnant Lid Parametrization Based on Temperature-Dependent Viscosity Laws

In recent years, experimental and numerical studies (e.g., Richter, 1978; Davaille and Jaupart, 1993; Solomatov, 1995; Moresi and Solomatov, 1995; Grasset and Parmentier, 1998; Reese *et al.*, 1999) have improved our understanding of convection in layers of fluids with strongly temperature-dependent viscosities. It has been shown that the simple scaling law (eqn [10]) is only applicable if the viscosity ratio across the layer is relatively small.

Three basic regimes of convection are found to occur, depending on the variation in viscosity between the hot interior of the convecting fluid and the cold surface: the small-viscosity regime, the transitional regime, and the stagnant lid regime. If the viscosity ratio across the layer is relatively small (smaller than about 100), the cold thermal boundary layer is mechanically similar to the rest of the fluid, and the convection planform and the temperature distribution are essentially the same as in isoviscous convection. For larger viscosity ratios (between about 100 and several thousand), the cold thermal boundary layer participates in the convection but deforms more slowly than the interior of the convecting fluid and thereby controls the heat transfer (transitional regime). With a viscosity ratio greater than about 10^3 , a so-called ‘stagnant lid’ will form on top of the underlying convecting layer. This stagnant lid (**Figure 3**) is not participating in the convective flow and most of the viscosity variation is within the lid.

Convection in the layer below the stagnant lid is driven by the nonadiabatic temperature difference across it. The viscosity variation in this convecting layer is small, that is, by only about one order of magnitude. Numerous studies assuming heating from below (e.g., Morris and Canright, 1984; Solomatov, 1995; Moresi and Solomatov, 1995) or

assuming heating from within (e.g., Grasset and Parmentier, 1998; Choblet and Sotin, 2000) find a viscosity ratio in the convecting layer smaller than 10. This observation is also independent of the assumed mantle rheology law, whether or not it is linear (Newtonian) or a power law (non-Newtonian).

All terrestrial planets other than the Earth presently show a stagnant lid on their surface. For such a lid to form, the following condition is required (Solomatov, 1995):

$$\theta = \frac{d \ln \eta}{dT} \Delta T > \ln 10 \quad [15]$$

For commonly assumed mantle rheology parameters, including, for example, the activation energy with $A \sim 400 \text{ kJ mol}^{-1}$ (e.g., Turcotte, 1982), and typical deep mantle temperatures of terrestrial planets (i.e., between about 1600 and 2500 K), the parameter θ is much larger than 10, which would imply that all terrestrial planets would be in the stagnant lid regime. A problem with the stagnant lid model, however, is that it may be ignoring some important aspects of mantle convection since the Earth also meets the above criterion for stagnant lid convection but instead has plate tectonics.

Davaille and Jaupart (1993) have shown with their experiments on internally heated fluids that the temperature at the base of the stagnant lid, T_l , can be expressed as a function of the temperature of the underlying convecting layer, T_m , measured just below the thermal boundary layer, and the rate of change of viscosity with temperature:

$$T_l = T_m - 2.21 \left(\frac{d \ln \eta}{dT} \right)^{-1} \quad [16]$$

Assuming a mantle viscosity as described by eqn [6] without pressure dependence, for stagnant lid convection, the temperature difference across the upper thermal boundary layer that drives the convection of the fluid (see Figure 3), if heated from within, is

$$\Delta T_{SL} = T_m - T_l = 2.21 \frac{RT_m^2}{A} \quad [17]$$

A similar relation has been obtained from numerical experiments by Grasset and Parmentier (1998) also with internally heated fluids. Choosing representative parameter values for a terrestrial mantle with $T_m \sim 1800 \text{ K}$ and $A \sim 400 \text{ kJ mol}^{-1}$, the difference between T_m and T_l is small, of the order of 100 K. This temperature difference across the thermal boundary layer is small in comparison to $(T_m - T_l)$ in earlier models based on constant-

viscosity laws. In these models, T_l equals the surface temperature and $T_m - T_l$ is about 1200 K or, if a rheological lithosphere is considered, T_l is about 1073 K (e.g., Schubert *et al.*, 1979b). In that case $T_m - T_l$ is about 700 K. Thus, the stagnant lid comprises the rheological lithosphere and is likely to be much thicker than the latter. As a consequence of the respectively high temperatures at the base of the stagnant lid, material in the lower part of the lid may have comparatively low viscosities and may be easily deformable on geological timescales.

In the present literature, there are at least two different approaches to parametrize the heat transport through a strongly temperature-dependent fluid in the stagnant lid regime. These two approaches differ in the time-dependent versus quasi-equilibrium treatment of the stagnant lid and are otherwise equivalent:

1. The approach by Grasset and Parmentier (1998) is based on the following rules for the stagnant lid regime: The interior of a layer adjusts so that the viscosity contrast across the convective thermal boundary layers is about 10. In the convecting layer, the scaling law for a fluid with constant viscosity is valid, and the Nusselt-Rayleigh number relationship for a fluid with constant viscosity is used (eqn [10]). The thickness of the stagnant lid is determined by equating the heat flux through the lid to the heat flux from the underlying convecting layer into the lid, according to eqn [14]. Instead of using an isotherm at the base of the stagnant lid, the temperature is time dependent and calculated from eqn [16]. The conductive heat transfer rate within the stagnant lid is calculated from a time-dependent heat-conduction equation.

2. In the second approach (e.g., Solomatov, 1995; Moresi and Solomatov, 1995; Reese *et al.*, 1999), a scaling law is derived for the whole layer, that is, the convecting part of the layer and the stagnant lid:

$$Nu = a\theta^c Ra_i^\beta \quad [18]$$

The constants a , c , and β vary with the properties of the convective regime. The internal Rayleigh number, Ra_i , is based on the total temperature drop across the layer and on the viscosity evaluated at the temperature of the interior below the stagnant lid and the upper thermal boundary layer. In addition to the internal Rayleigh number a second parameter is introduced for this approach, the Frank-Kamenetskii parameter, θ , which is the natural logarithm of the viscosity contrast across the convecting layer (eqn [15]).

10.09.4.3 Plate Tectonics, Lithosphere Delamination, and Stagnant Lid Convection

The parametrizations based on constant and temperature-dependent viscosity laws introduced in Sections 10.09.4.1 and 10.09.4.2 are used to describe the heat

transport mechanisms in a terrestrial planet by either plate tectonics, lithosphere delamination, or stagnant lid convection (Figures 3 and 4). In the following, we use for all model calculations parameters that correspond to the planet Mars, see Table 1).

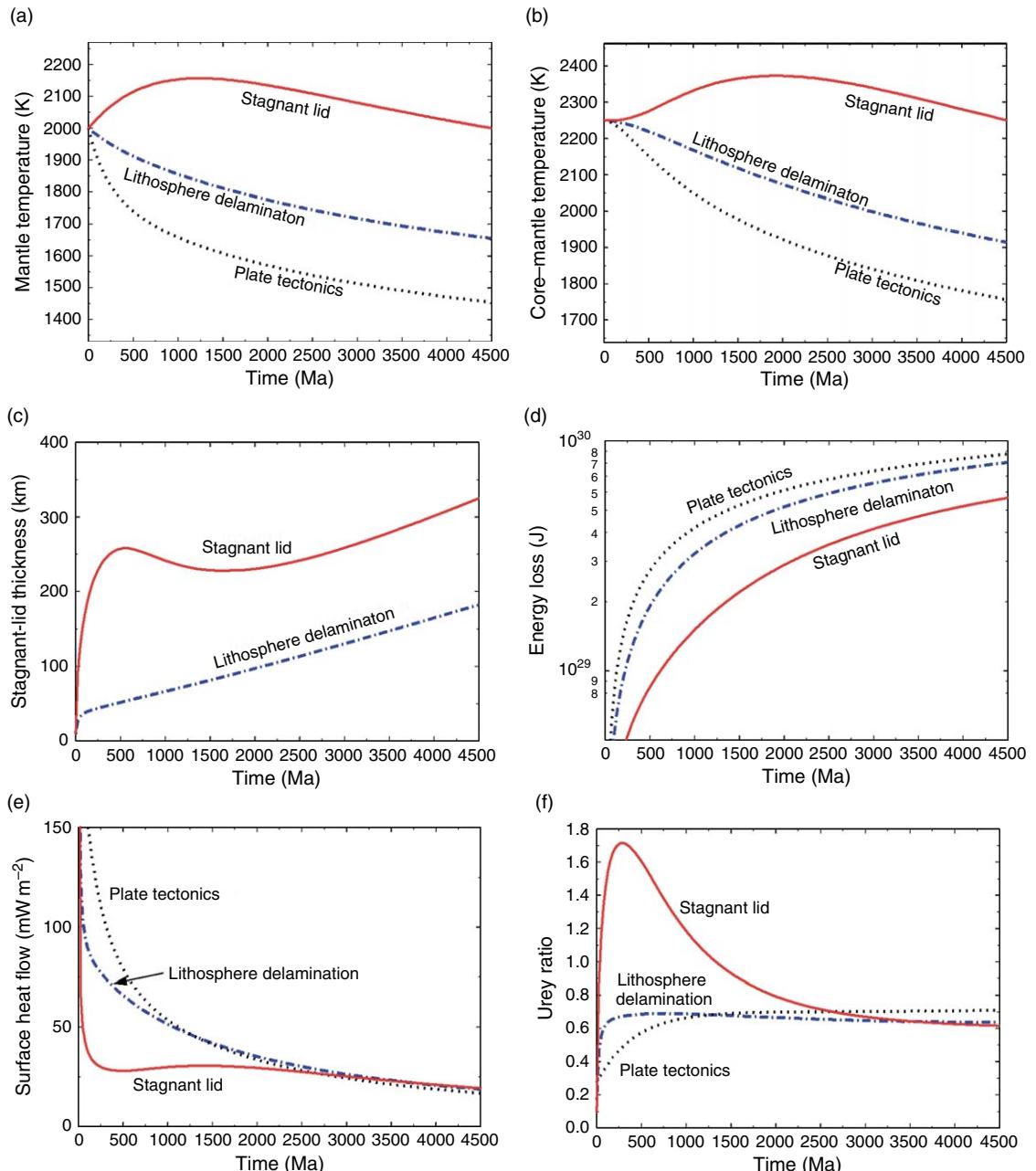


Figure 4 Mantle temperature (a), core–mantle temperature (b), stagnant lid thickness (c), energy loss (d), surface heat flow (e), and Urey ratio (f) as a function of time for three different heat transfer models: stagnant lid convection (red line) and lithosphere delamination (dash-dotted blue line), and plate tectonics (dotted line). For further description see text.

Table 1 Definition of parameters and their values for the thermal evolution models in Section 10.09.4.3

Parameter	Notation	Value	Unit
Radius of planet	R_p	3400×10^3	m
Radius of core	R_c	1700×10^3	m
Gravity	g	3.7	m s^{-2}
Density of mantle	ρ_m	3500	kg m^{-3}
Heat capacity of mantle	C_m	1000	$\text{J kg}^{-1}\text{K}^{-1}$
Mantle thermal expansivity	α	2×10^{-5}	K^{-1}
Mantle thermal diffusivity	κ	10^{-6}	m^2s^{-1}
Mantle thermal conductivity	k	4	$\text{W m}^{-1}\text{K}^{-1}$
Surface temperature	T_0	220	K
Gas constant	R	8.314 4	$\text{J mol}^{-1}\text{K}^{-1}$
Initial heat source density	Q_0	1.6×10^{-8}	W m^{-3}
Decay constant	λ	10^{-17}	s^{-1}
Exponent of the Nu-Ra relation	β	1/3	
Critical Rayleigh number	Ra_{crit}	450	

To calculate the thermal evolution, the energy balance of the mantle and the core has been solved.

$$\rho_m C_m V_m \frac{dT_m}{dt} = -q_m A_m + Q_m V_m \quad [19]$$

$$\rho_c C_c V_c \frac{dT_c}{dt} = -q_c A_c \quad [20]$$

where V_m and V_c are the volume of the convecting mantle and the core, A_m and A_c are the surface area of the mantle and the core, T_m and T_c are the temperature of the isothermal mantle below the conductive layer and the temperature of the core, respectively, q_m and q_c are the heat flow out of the mantle and the core, and Q_m is the heat production rate in the mantle

$$Q_m = Q_0 \exp(-\lambda t) \quad [21]$$

where Q_0 is the initial heat-generation rate, λ is the decay constant, and t is the time.

The heat flow of the mantle and the core are calculated from eqn [12] with δ the thickness of the upper or lower thermal boundary layer, and ΔT_δ the temperature difference across the upper or lower boundary layer. The thickness of the boundary layers is calculated from eqn [13] replacing d with δ and assuming that the layers become unstable when Ra reaches a critical value, Ra_{crit} . The lid of the stagnant lid and the lithosphere delamination model grows in thickness with time according to eqn [14].

Stagnant lid parametrization models show that the lid will thicken rapidly as the planet cools while the temperature of the underlying convecting mantle and core will change comparatively little (Figures 4(a)–4(c)). With plate tectonics, there is, of course, no growth of a lid and the cooling of the deep interior is most efficient. The evolution of the lid thickness and the mantle temperature for the lithosphere

delamination model is in between the two extremes. For the stagnant lid model, the conductive layer grows rapidly during the first few hundred million years and its thickness varies comparatively little thereafter. Plate tectonics and lithosphere delamination models show an approximately constant rate of increase in the conductive-layer thickness. Although the present-day thickness of the stagnant lid differs significantly among the models (320 km for the stagnant lid model, 180 km for the lithosphere delamination model, and no lid for the plate tectonics model) the present-day thickness of the conductive layer is relatively similar for all three models.

Figures 4(d) and 4(e) compare the cumulative energy loss, that is, the energy loss over the entire evolution of the planet, and the surface heat flow of the three models with $A = 300 \text{ kJ mol}^{-1}$. Stagnant lid convection cools the entire planet least effectively. After 4.5 Ga, the energy loss is about $3.1 \times 10^{29} \text{ J}$ smaller for the stagnant lid model compared to the plate tectonic model and $2.3 \times 10^{29} \text{ J}$ smaller for the stagnant lid model compared to the lithosphere delamination model. Such energy losses, when distributed homogeneously over the entire mantle, imply that the present-day mantle is about 600 K cooler in the plate tectonics model and about 450 K cooler in the lithosphere delamination model than it is in the stagnant lid model.

It may be speculated that a planet, in which heat is transported effectively by plate tectonics, has a higher surface heat flow than a planet in the stagnant lid regime. However, this is only true in the early evolution. After about 2.5 Ga, the surface heat flow is even slightly higher in the stagnant lid model. The albeit small difference in the present-day surface heat

flow has also implications for the Urey ratio, that is, the ratio of the heat produced within the mantle by radioactive decay to the total surface heat flow (**Figure 4(f)**). The present-day Urey ratio for our Mars-like test planet is about 0.6 for the stagnant lid regime and 0.7 with plate tectonics. For the lithosphere delamination model, the present-day value is in between the two models of stagnant lid convection and plate tectonics. The contribution of secular cooling at present is, therefore, slightly higher for a planet in the stagnant lid regime than for a planet with plate tectonics assuming the same mantle parameters.

A strong decrease of the surface heat flow can be observed during the first 250 Ma for the stagnant lid model. The strong decrease in the surface heat flow is caused by the adjustment of the lithosphere, which grows rapidly from 10 to 250 km as a consequence of the inefficient heat flow out of the convecting mantle and the more efficient heat transport through the conductive layer. The thicker the assumed initial lithosphere and the lower the initial mantle temperatures, the slower is the decrease of the surface heat flow in the early evolution. The cooling behavior during the early stage of the planet's evolution is, therefore, strongly dependent on the initial conditions. It has been argued by Choblet and Sotin (2000) that an initial stage up to 1 Ga cannot be described by the stagnant lid parametrization, which is based on steady-state convection models. During that initial transitional stage the mantle cools by conduction until the upper thermal boundary layer becomes unstable due to increased mantle temperatures and reduced viscosities, and convection sets in.

10.09.5 Thermal Evolution, Volcanic History, and Magnetic Field History of Terrestrial Planets

The various parametrized approaches, described above, have been used to calculate the thermal histories of the terrestrial planets and a significant amount of work has been published to date (e.g., Sharpe and Peltier, 1978; Schubert 1979; Schubert *et al.*, 1979a; Stevenson *et al.*, 1983; Schubert *et al.*, 1990; Spohn, 1991; Schubert *et al.*, 1992; Hauck and Phillips, 2002; Breuer and Spohn, 2003). According to these previous models, the terrestrial planets are differentiated rapidly into a solid mantle and a fluid core (see also Sections 10.09.2.1 and 10.09.2.2). The temperature distribution subsequent to core formation, therefore, serves as the initial boundary

condition for the thermal evolution models. Early thermal evolution models use fixed initial temperatures (e.g., Schubert and Spohn, 1990; Spohn, 1991; Schubert *et al.*, 1992; Breuer *et al.*, 1993). The main reason not to vary the initial temperature distribution, which is basically unknown, can be justified by a characteristic feature of the thermal evolution models: The strong temperature dependence of mantle viscosity acts as a thermostat regulating the average mantle temperature (Tozer, 1965). For a hot planet, mantle viscosity is low, and extremely vigorous convection rapidly cools the planet. For a relatively cool planet, the mantle viscosity is higher, and more moderate convection cools the planet at a reduced rate. Self-regulation tends to quickly bring the viscosity of the mantle to a value that facilitates efficient removal by convection of the heat generated in the mantle. Even an initially cold mantle would heat up by radioactivity until the self-regulated viscosity is reached. As a consequence of the self-regulation, the present state of the convecting mantle has little or no memory of the initial conditions. This circumstance makes the early thermal evolution models more applicable, in particular because the initial state is unknown. However, if one is interested in the crust growth and the magnetic field generation of a planet, in particular in its early evolution, knowledge about the early temperatures distribution is required (e.g., Hauck and Phillips, 2002; Breuer and Spohn 2003, Breuer and Spohn, 2006).

Crustal growth and dynamo action in the core are strongly coupled with the thermal evolution of a planet since both processes depend on the vigor and the mode of convection in the mantle. A model of a planet's evolution and tectonics is therefore required to explain the evolution of the magnetic field and it must also account for the differentiation of the planet as well as for the formation and growth of its crust. Furthermore, the thickness of the elastic lithosphere places important constraints on the thermal evolution of a planet. The elastic lithosphere is the upper layer of a terrestrial planet that supports stresses over geologically long ($\sim 10^8$ years) intervals and can be derived by a joint analysis of the gravity and topography data or from flexural modeling to match the topography (see Chapter 10.05). Since the base of the elastic (mechanical) lithosphere can be defined by an isotherm which depends on the rheology (composition) and the strain rate (McNutt, 1984), its thickness allows estimates of the surface heat flow. It is important to note that the elastic lithosphere thickness obtained by gravity and topography data, does not

represent the current state of the thermal lithosphere. It rather shows the thermal state at the time of formation of the considered geological structure. Thus, the elastic lithosphere thickness provides constraints on models of thermal evolution.

For a typical terrestrial mantle material dominated by olivine this isotherm can be estimated to be about 1050 K using the rheology parameters of Karato *et al.* (1986) and a strain rate of 10^{-17} s^{-1} . This isotherm defines the maximum thickness of the elastic lithosphere since the thickness of the effective elastic lithosphere decreases with its deflection. Furthermore, the crust can significantly reduce the elastic thickness since the crust is intrinsically weaker than the mantle (e.g., Burov and Diamant, 1995).

10.09.5.1 Crust Formation

The crust of a terrestrial planet forms when partial melt of the mantle rises toward the surface. Thus, the composition, the thickness, and the time of emplacement of the crust can give us important constraints on the thermochemical evolution; it tells us about the interior temperature distribution and how the processes forming the crust vary with time. On the other hand, the crust itself influences the thermal evolution. (1) Melt that forms the crust is enriched in radioactive elements. The consequence is the redistribution of the radioactive elements and in particular a depletion of the mantle in these elements. (2) Crustal material has a different composition and structure than mantle material. The consequence is a lower thermal conductivity of the crust in comparison to the conductivity of the mantle – thus, a less efficient heat transport through the outer layer.

10.09.5.1.1 Primary and secondary crust formation

The process of crust formation on terrestrial planets can, in general, be divided into two phases: primary and secondary crust formation. A primary crust can be formed as a consequence of efficient differentiation due to core formation and/or freezing of a magma ocean. Such a primordial crust can be strongly enriched in radioactive elements as suggested for the anorthositic crust on the Moon and may have strongly depleted the mantle (for a review, see Warren (1985)). Except for the Moon, the existence, persistence, and composition of such a crust for the terrestrial planets is however unknown.

The presence of a magma ocean on a terrestrial planet suggests chemical stratification of the mantle due to fractional crystallization. Fractional crystallization of a cooling magma ocean is generally accepted as an explanation of the observed dichotomy between Eu-enriched anorthositic crust, and the Eu-depleted basaltic source regions of the Moon (e.g., Wood, 1970). Assuming the magma ocean freezes rapidly during about 100–200 Ma (e.g., Solomon and Longhi, 1977; Minear and Fletcher, 1978) without any disturbances, this part of the mantle most likely became chemically stratified following magma ocean differentiation and crystallization with the late dense iron-rich phases coming to rest upon a less-dense Mg-rich phase. Furthermore, most of the incompatible and heat-producing elements would have been enriched in a KREEP-rich layer that is located just below the anorthositic crust leaving the mantle of the original magma ocean strongly depleted in radioactive heat sources. Chemical fractionation, which possibly results in mantle layers of different compositions and densities, most likely would have strongly influenced the subsequent thermochemical evolution of the planet (e.g., Hess and Parmentier, 1995; Elkins-Tanton *et al.*, 2003). However, for the larger terrestrial planets like Earth and Venus, and possibly for Mars, Tonks and Melosh (1990) argue that the higher pressure gradient for those planets as compared to the Moon and Mercury inhibits efficient chemical fractionation. The magma ocean rather would have started to cool via equilibrium crystallization. The existence of a magma ocean in general has been questioned from numerical modeling of a growing Mars (Senshu *et al.*, 2002). The models show no global magma ocean formed during accretion; instead small magma ponds form by the impacts.

In the subsequent evolution after freezing of the magma ocean, a secondary crust that is usually basaltic is formed by partial melting of the mantle. Depending on the degree of early mantle differentiation, the entire mantle (or only part of the upper mantle) is already depleted in crust-forming material. The formation of the secondary crust depletes the mantle further in radioactive elements as those are incompatible and enriched in the melt that form the crust. The mechanisms of crust formation differ between a plate tectonics and a one-plate planet.

Plate tectonics provides for crust formation in a two-step fashion (see Condie, 1997) for a discussion of plate tectonics and the Earth's crust and Schubert *et al.* (2001) for a discussion of plate tectonics and

mantle convection). Crust formation is most effective at divergent plate boundaries, where rising hot mantle material crosses the solidus near the surface. This pressure-release melting generates basaltic crust that is continuously recycled at convergent plate boundaries. More silicic crust is produced in a second differentiation step at convergent plate boundaries where basaltic crust is re-melted (together with continental sediments and possibly mantle rock) to form new continental-like crust.

For one-plate planets, there is no crust recycling and there is no two-stage differentiation. Instead, melt is formed underneath the lithosphere usually at greater depth than with plate tectonics. The crust produced is assumed to be basaltic. In contrast to crust formation at the divergent plate boundaries on plate tectonic planets, crust growth is limited by the increasing thickness of the lithosphere as the planet cools. First, because buoyancy of the melt decreases with increasing depth to the source region (the melt is more compressible than the mantle rock). Second, the melt needs to rise through the ductile lower and the upper elastic part of the lithosphere. For the latter, the permeability of the elastic part of the lithosphere to magma ascent decreases with increasing thickness and pressure. Nonetheless, volcanism is observed even on a body with a thick crust and weak gravity such as the Moon.

Secondary crust formation may also result in a chemical layering of the mantle as suggested for the primary crust that forms from freezing of a magma ocean. When basaltic melt is extracted from the mantle and segregates to form the surface layer, fertile mantle rock reduces to harzburgite. Under certain melting conditions (Schutt and Lesher, 2006), harzburgite is less dense than its fertile source rock, this layer can have a stabilizing effect on the mantle dynamics and prevent further cooling of the planet. Depending on the thickness of the depleted layer, basically three different evolution scenarios are possible: (1) the layer is thinner than the stagnant lid and thus has no influence on the thermal evolution of the planet; (2) the harzburgite layer grows faster than the stagnant lid, further slowing the cooling of the planet as the thickness of the nonconvecting part of the mantle increases and inefficiently transports heat by conduction; (3) the harzburgite layer is sufficiently thick that it can convect separately from the lower fertile mantle. Such two-layer convection results in inefficient cooling of the deep interior. The thickness of the depleted harzburgite layer depends on the mantle temperature and especially

on the depth of melting, that is, for high temperature the melting is deeper, and the depleted layer is thicker.

10.09.5.1.2 Plume volcanism and partial melt zone

As mentioned above, in one-plate planets melt is usually generated underneath the lithosphere in places where the mantle temperature is higher than the mantle solidus. The melt can be either produced locally due to upwelling hot material, that is, mantle plumes, or more globally in a partial melt zone if the average mantle temperature is higher than the solidus.

Whether a plume and associated volcanism can occur depends on the existence of a thermal boundary layer at the base of the mantle or at an internal chemical interface. Plumes are initiated in the boundary layers and are sustained by thermal instabilities. The existence of a thermal boundary layer at the bottom of the mantle requires sufficient heat flow from the core. If the mantle is mainly heated from within, plumes are very weak and the associated volcanism is either insignificant or not present. In this case, mantle convection is dominated by cold downwellings and the upwellings are broad with a small temperature contrast relative to the average mantle. Mantle convection and thermal evolution models show that planetary mantles of one-plate planets are mainly dominated by internal heating during a long period of their evolution. An exception can occur in the very early evolution if the core is superheated due to core formation or if sufficient heat is generated in the core due to the decay of radioactive heat sources. For example, potassium might be incorporated in the core at high pressure (Murthy *et al.*, 2003). In case of a superheated core, heat flow from the core decreases rapidly a few hundred million years after core formation and the temperature difference across the CMB layer decreases rapidly to only a few tens of kelvins (Spohn *et al.*, 2001; Hauck and Phillips, 2002; Breuer and Spohn, 2003, 2006). Convection models show that plumes cease or cannot even develop under these conditions (Spohn *et al.*, 2001). For example in Mars, the temperature difference at the CMB, which is the maximum excess temperature of the plume, can be increased to larger values of about 150 K if the viscosity increases with depth (Buske, 2006). The excess temperature further decreases with height above the CMB because of adiabatic cooling and heat diffusion (Albers and Christensen, 1996) and in

the presence of the exothermic phase transitions, for example, α - to β -olivine and β - to γ -olivine. In conclusion, one would expect only weak volcanic activity due to present-day plumes in a planet with one-plate tectonics. However, present plume volcanism is possible for the following reasons: (1) in case of a layered mantle due to a chemical interface or an endothermic phase transition, a thermal boundary layer can form at the interface of the convecting mantle layers where upwelling plumes may develop; (2) a strong increase of viscosity with depth; and (3) heat supply from the core due to radioactive elements. The maximum excess temperature of these plumes, that is, the temperature jump across the thermal boundary, however, depends strongly on the heat-source distribution within the mantle and the core, the mantle rheology, and the existence of phase transitions.

In contrast to strong mantle plumes, global partial melt zones can be present for a much longer time period, probably until the present, depending on the efficiency of mantle cooling and amount of radioactive heat sources. For example, a one-plate planet with a dry mantle, that is, a mantle with a stiff rheology, a thick low-conducting crust, and a high concentration of radioactive elements in general shows a persistent partial melt zone. A typical evolution of such a melt zone involves the downward propagation of its upper boundary with time as a consequence of planetary cooling. Thus, the average source region of the melt migrates to greater depth. However, there have been no studies of how a global partial melt zone influences the efficiency of the heat transport.

10.09.5.1.3 Thermal properties of the crust

The crusts of terrestrial planets can act as thermal insulators for the planets. Crust is enriched in radioactive elements and its thermal conductivity is lower than that of mantle material. The thermal conductivity of most compact volcanic materials ranges between 1.5 and $3.5\text{ W m}^{-1}\text{ K}^{-1}$ at ambient temperatures (Clifford and Fanale, 1985; Clauer and Huenges, 1995) but it decreases with temperature similar to mantle material (Seipold, 1998). For example, the thermal conductivity of typical compact basalt decreases from about $2\text{ W m}^{-1}\text{ K}^{-1}$ at 270 K to about $1.5\text{ W m}^{-1}\text{ K}^{-1}$ at 800 K . In addition to the temperature effects on thermal conductivity, the structure of the material can significantly change the thermal conductivity. Fractured and porous

materials have a reduced thermal conductivity in comparison to compact material. The upper crust of the terrestrial planets and the Moon has been fractured due to impact processes, in particular in the early period of heavy bombardment. These impacts resulted in the production of a porous megaregolith that extends to considerable depth, for example, to about 10 km for Mars (Clifford, 1993) and about 20 km for the Moon (Binder and Lange, 1980). At the transition between fractured and coherent basement, the lithostatic pressure is sufficient to close all fractures and the intergranular pore space. It is expected that the porosity decreases exponentially from the dusty surface to this self-compaction depth (Binder and Lange, 1980). As a consequence, thermal conductivity decreases from the bottom of the megaregolith toward the surface. At low atmospheric pressures (e.g., around the ambient atmosphere pressure on Mars and lower), surface thermal conductivity can take on extremely low values. Remote thermal measurements for the Martian surface indicate soil thermal conductivities in the range of 0.075 – $0.11\text{ W m}^{-1}\text{ K}^{-1}$ (Kieffer, 1976) but it is assumed that thermal conductivity increases significantly in the first few centimeters, for example, to a value of about $0.5\text{ W m}^{-1}\text{ K}^{-1}$ (Helbert and Benkhoff, 2003), typical of sedimentary material or loose sandstone. The decrease of thermal conductivity due to porosity is less if the pores are filled, for example, with water ice. Furthermore, hydrothermal circulation in a porous, permeable regolith can counteract the low conductivity of the crust. This possibility might work for Mars since there are indications that the planet had a water-rich crust for much of its geological history.

10.09.5.1.4 Observations

To study the crustal evolution and composition of the terrestrial planets, a combination of global spectral imaging, optical imaging, gravity measurements, and laser altimetry (remote sensing), *in situ* investigations, and laboratory analyses of returned samples is desired (cf. Chapter 10.18). Since there is a different level of planetary exploration for each of the terrestrial planets, our knowledge varies strongly among the planets with Mars and the Moon being the most studied bodies so far.

The mapping of the surfaces using image and topography data allows us to reconstruct geological units. Of particular importance is the determination of the relative ages of different units by crater counting. In the case of Venus, the dense atmosphere does

not allow us to obtain visible images of the surface but with the radar on board the NASA Magellan mission it was possible to study the surface with visible-like radar images and topography data thereby allowing crater studies.

To obtain the absolute ages of the geological units, direct samples from the planetary surfaces are required. These dated samples are available from the Moon but not for other terrestrial planets (SNC meteorites that are believed to originate from Mars cannot be used due to lack of knowledge of the geologic setting from where the meteorites came). Therefore, the cratering rate and the absolute ages of the Moon's surface are used to calibrate the relative surface ages of the other terrestrial planets. However, there is an uncertainty in the ratio of the specific planet/Moon cratering rates which propagates directly into uncertainty on age (*see Chapter 10.06*). Despite this uncertainty, with the identification of volcanic areas and their ages, the image and topography data can also be used to estimate the amount of volcanic activity with time. However, one needs to consider that only extrusive volcanism can be observed at the surface but not intrusive material below the surface. Estimates of the relative proportion between intrusive and extrusive material range from a factor of 7 to 10 (Greeley and Schneid, 1991), but this factor might vary strongly among the planets and the crustal formation processes. Further insight to the timing of crustal evolution can be obtained by the geochemical analysis of samples. Such studies exist for the Moon (Apollo samples) and Mars (SNC meteorites) that have identified for both bodies an early mantle differentiation event with a major crust formation event 4.5 Ga.

Seismology is the method of choice for determining crustal thickness on Earth. Seismic data are not yet available for the other terrestrial bodies (except lunar seismic data acquired during the Apollo exploration), although the deployment of seismometers on Mars is envisioned for future missions. In the absence of seismological data, however, the crustal thickness of terrestrial planets and the Moon is nonuniquely constrained by the gravity field and topography measurements (*see Chapter 10.05*).

The composition of the crust can be obtained by a combination of different methods. The most reliable method is the examination of samples in the laboratory, possible so far for the Moon and Mars. Without direct access to samples, the mineralogical composition can also be assessed on a global scale by infrared (IR) imaging spectroscopy in the range of 0.8–50 μm.

This spectral domain contains a large number of signatures diagnostic of many minerals. For *in situ* measurements, the alpha-proton X-ray spectrometer (APXS) technique has been used on the NASA Pathfinder mission to Mars and the Mars Exploration Rovers (MER), Spirit and Opportunity, to chemically analyze soils and rocks without any sample preparation (except the use of a rock abrasion tool (RAT) to analyze 'fresh' crustal material). The elemental composition, however, is not as diagnostic as mineralogical composition as a variety of minerals might have similar chemical composition. Other *in situ* instruments are the Mössbauer spectrometer to determine with high accuracy the composition and abundance of iron-bearing minerals and the mini-thermal emission spectrometer to analyze the mineral composition of Martian surface features and select specific rocks and soils for more detailed investigations. Both instruments have been already used on the MER rovers.

10.09.5.1.5 Incorporation of crustal growth in thermochemical evolution models

To calculate crustal growth and mantle differentiation in thermochemical evolution models, one usually assumes that both effects are regulated by the mantle convection speed (e.g., Turcotte and Huang 1990; Schubert *et al.*, 1990; Hauck and Phillips, 2002; Breuer and Spohn, 2003). In earlier models, crustal growth was simply proportional to the vigor of mantle convection and a fixed parameter measuring the efficiency of magma generation (Schubert *et al.*, 1990; Spohn 1991). These models show crustal production rates even when the mantle does not contain partial melt. In more recent parametrizations, the actual melt fraction as a function of the melting and mantle temperature is considered in more detail (Hauck and Phillips, 2002; Breuer and Spohn, 2003, 2006). The parametrizations in the literature are slightly different, for example, in how they incorporate the increase of the melting temperature due to continuous depletion of the mantle in crust-forming elements. However, they basically give similar results (Hauck and Phillips, 2002; Breuer and Spohn, 2003, 2006).

10.09.5.2 Magnetic Field Generation

Magnetic field generation in the iron core of a terrestrial planet (*see Chapters 10.08 and 8.03*) is strongly coupled to the thermal evolution of the planet. Thus,

knowledge about the magnetic field evolution can be used to constrain the thermal evolution of a planet. Planetary space missions instrumented with magnetometers provide insight about the present magnetic fields of the terrestrial planets. The available data, however, differ significantly (see Chapter 10.07). In the case of the Moon, the lunar magnetic field has been studied indirectly via the natural remanent magnetization of the returned lunar samples, and directly with magnetometers carried to the surface and placed in orbit at low altitude above the surface on the Apollo 15 and Apollo 16 subsatellites (see reviews in Dyal *et al.* (1974) and Fuller and Cisowski (1987)). In 1998, Lunar Prospector allowed detailed global measurements of the magnetic field (Lin *et al.*, 1998). In the case of Mars, Mars Global Surveyor (MGS) measured the magnetic field of the entire planet (Acuña *et al.*, 1998). Both bodies, that is, the Moon and Mars, show no sign of a present internal field but parts of their crusts preserve a remanent magnetization (Coleman and Russell, 1977; Acuña *et al.*, 1999). It is assumed that the remanent magnetization was established at the time of crust formation while an internal dynamo was active (Runcorn 1975; Acuña *et al.*, 1999). An age determination of the magnetized and nonmagnetized crust would thus allow us to reconstruct the magnetic field history.

In the case of Mercury, we have only two flybys by Mariner 10 that identified a weak internal field (Ness *et al.*, 1974). Venus, like Mars and the Moon, has no present-day dynamo action in its core. For both planets, Mercury and Venus, we cannot exclude the presence of an old magnetized crust because data are not available so far, but the temperature of the crust of Venus is likely above the Curie point even at the surface. As long as we do not know whether the crust is magnetized, we are unable to reconstruct the magnetic field history. We only know the present state and information about the early magnetic field evolution is only possible by modeling.

10.09.5.2.1 Core convection and necessary conditions for dynamos

To generate a self-sustained magnetic field, core convection is required. To obtain core convection, one must appeal to unusually large heat flows or the development of an inner core. In either case, the core must be cooling and the cooling is controlled by the heat transport of the outer layers, that is, the mantle and the crust.

10.09.5.2.1.(i) Thermal dynamo Thermal convection in the core, like thermal convection in the mantle, is driven by a sufficiently large super-adiabatic

temperature difference between the core and the mantle. It occurs if the core heat flow exceeds that conducted along the core adiabat. The latter heat flow, therefore, serves as a criterion for the existence of thermally driven convection in the core.

$$q_{\text{crit}} = k_c \frac{dT}{dr} = k_c \frac{dT}{dP} \frac{dP}{dr} = k_c \frac{\alpha_c g_c T}{C_c} \quad [22]$$

with k_c the thermal conductivity of the core, dT/dr the temperature gradient in the core, P the pressure, α_c the thermal expansivity, C_c the heat capacity, and g_c the gravity. The critical heat flow varies between 5 and 20 mW m⁻² for Mars (Nimmo and Stevenson, 2000) and Mercury, between 15 and 40 mW m⁻² for Venus, and is smaller than 7 mW m⁻² for the Moon. The large variation in the value for each planet comes partly from the uncertainty in the thermal conductivity of the core material, which varies between 43 and 88 W m⁻¹ K⁻¹ at the core pressures (Anderson, 1998). The cooling of the core is strongly controlled by the thermal evolution of the overlying mantle. If the mantle removes heat from the core at a rate that exceeds the critical heat flow, then the core will convect. If the mantle removes heat at a rate less than the critical heat flow, the core is thermally stably stratified and dynamo action by thermal convection is not possible.

10.09.5.2.1.(ii) Chemical dynamo Compositional convection can occur due to the release of positively buoyant material during the process of solid inner-core freezing from a fluid core with noneutectic composition (Braginsky, 1964). The existence of light elements in the core has been assumed for all terrestrial planets (see Chapter 10.02) and many suggestions have been offered for the amount and identity of the light elements that are mixed with the iron. Chemical convection can drive a dynamo more effectively than thermal convection (Braginsky, 1964; Stevenson *et al.*, 1983). In contrast to the transformation of thermal energy to magnetic field energy, where only a small part of the energy is transferred (described by the Carnot efficiency factor, which is between 0.06 and 0.11; Stevenson *et al.*, 1983; Lister and Buffett 1995), a chemically driven dynamo has no similarly reduced efficiency.

Chemical convection and the associated generation of a magnetic field in the core occur if the temperature in the fluid (outer) core lies between the solidus and the liquidus of the core material. Inner-core growth permits outer-core convection even when the heat flow through the CMB is less

than the heat carried by conduction along the adiabat.

In conclusion, whether a terrestrial planet has core convection and whether it can have a thermal or chemical dynamo depends on the cooling rate of the core and the core temperatures. If the mantle transports sufficient heat, then a dynamo occurs, but if the mantle is too insulating, then there is no dynamo.

Thermal evolution models indicate that the existence of a present-day magnetic field for a terrestrial planet is associated with the growth of an inner core. A present-day thermal dynamo is unlikely since the results show a slow cooling of a core with heat flow less than the heat flow along the core adiabat. In fact, a thermal dynamo is in general difficult to generate and only likely if the core is superheated with respect to the mantle due to the core-formation process. If a thermal dynamo exists, it shuts off very early in the evolution as the heat flow out of the core decreases rapidly during the first few hundred million years. Whether in the subsequent evolution an inner core can grow and thus a compositional dynamo can be initiated depends strongly on the temperature evolution and the core composition.

10.09.6 Comparison of the Terrestrial Planets and the Moon

Although the evolution, composition, and interiors of terrestrial planets have some features in common (such as iron-rich cores, slowly convecting silicate mantles, and rigid crustal layers), they are entirely different in many other respects.

For instance, a strong diversity in crustal evolution has been identified: Mercury and the Moon show highly cratered surfaces indicating an average crust older than about 4 Ga with only minor volcanic activity in the subsequent evolution. Mars shows a distinct dichotomy of the surface consisting of the old southern highlands with an average age of more than 4 Ga and the superficially younger northern lowlands with an average age of about 3.5 Ga (Tanaka *et al.*, 1992). Below the comparative young surface, the northern crust has a similar age as the southern crust as concluded from topography data of partly buried impact structures (Frey *et al.*, 2002). A continuous and rapid decline of global volcanism and crustal growth since planet formation can be observed in the surface morphology. Over time, global volcanism has been concentrated into mainly two

volcanic regions, Tharsis and Elysium. In comparison to the other terrestrial planets, except the Earth, the average surface of Venus seems to be the youngest with an average crustal age of 300–800 Ma (Schaber *et al.*, 1992; McKinnon *et al.*, 1997).

Similar to crustal evolution, magnetic field evolution differs among the terrestrial planets and the Moon. Mercury is the only terrestrial planet, except the Earth, that has a magnetic field (Ness *et al.*, 1974) – although a weak one – that is generated in the iron core (Jupiter's satellite Ganymede also shows an internally generated field; Kivelson *et al.*, 1996). Venus, Mars, and the Moon on the other hand have no present global magnetic fields but at least in the case of the Moon and Mars the remanent magnetization of part of their crust indicates that they once had dynamo action in their cores (Runcorn 1975; Acuña *et al.*, 1999). The diversity in the volcanic and magnetic history of the bodies is also reflected in their thermochemical evolution. In the following sections, we will describe the thermal, volcanic, and magnetic evolution of the planets.

10.09.6.1 Mercury

During its three flybys of Mercury, the Mariner 10 spacecraft returned images of about 45% of the planet's surface. The images revealed a lunar-like terrain with an average age of more than 4 Ga and lobate scarps that are more or less evenly distributed over the well-imaged portion of the surface. The scarps are thought to indicate an average contraction of the planet's radius by 1–2 km since the end of heavy bombardment (Strom *et al.*, 1975) and provide an important constraint on the planet's thermal evolution. Whether this constraint provides an upper bound on global contraction, however, depends on whether strain has been accommodated by other mechanisms, such as faults unresolved in Mariner 10 images (Strom *et al.*, 1975) or long-wavelength, low-amplitude folds (Dombard *et al.*, 2001). The most likely source of global contraction is a combination of a thermal contraction (reduction in average internal temperature) and a phase change by solidification (e.g., growth of an inner core). Core freezing contributes more to the global contraction. Complete core solidification would shrink the planet about 17 km in radius (Solomon, 1976), much more than required from the geological observations. A small inner core is therefore better to reconcile with a small planetary contraction.

Thermal evolution models (Schubert *et al.*, 1988; Hauck *et al.*, 2004) indicate that the observed scarps account for only a rather small fraction of the total contraction. One possible explanation is that much of the contraction would have predated the observable geologic record – for instance, most of the inner core was formed before the end of heavy bombardment. In any case, to meet the constraint of only 1–2 km of contraction since the end of heavy bombardment, relatively little cooling of the interior should have happened since that time.

The thermal history of terrestrial planets depends strongly on the amount of radioactive elements in its interior. Predictions for the abundances of the important heat-producing elements uranium, thorium, and potassium in Mercury depend on the formation scenario, which seems to be very special among the terrestrial planets, as indicated by the Mercury's large bulk density of $\sim 5430 \text{ kg m}^{-3}$. There are three general hypotheses for the formation of Mercury. The first scenario for forming Mercury's large bulk density involves aerodynamic sorting of iron and silicate particles in the solar nebula (Weidenschilling, 1978) that might have given the planet abundances of U and Th similar to the upper mantle of the Earth, though lacking much K (Basaltic Volcanism Study Project, 1981). The other two models for Mercury's formation involve a later-stage loss of silicate, and retention of iron, materials. Impact-induced stripping of the outer, silicate layers of a larger, differentiated, proto-Mercury (e.g., Cameron, *et al.*, 1988; Wetherill, 1988) could have left a silicate layer with near-CI chondritic abundances of heat-producing elements (e.g., Lodders and Fegley, 1998). A final, more exotic scenario, involves a late-state vaporization of silicates that would result in a Th-rich silicate layer lacking appreciable U and K (Fegley and Cameron, 1987). Of the three formation models, the vaporization model predicts less cooling and contraction of the planet (Hauck *et al.*, 2004) as it assumes ^{232}Th as the sole of radiogenic heat (Fegley and Cameron, 1987), which with its 14 Gyr half-life has not had a significant decline in heat output.

Due to the specific interior structure of Mercury, that is, the relatively thin mantle, a convecting mantle cools the interior efficiently and it has been speculated that a pure iron core could be totally frozen very early in the evolution (Siegfried and Solomon, 1974; Solomon, 1976). Such a scenario is consistent with the geological observation of small global contraction since the end of the heavy bombardment. A totally frozen core, however, is difficult

to reconcile with the magnetic field history of Mercury. During the first and third close encounters in 1974 and 1975, Mariner 10 passed briefly through a small but earth-like magnetosphere. The analyses of these observations revealed the presence of an internal field, with a dipole moment that is a factor of about 10^4 smaller than that of the Earth (Ness *et al.*, 1974, 1975, 1976; Russell *et al.*, 1988; see review in Connerney and Ness, 1988). The discovery of the internal field strongly suggests the presence of a fluid outer core. To prevent the core from freezing, the addition of radioactive heat sources into the core (Toksöz *et al.*, 1978), a late core formation (Sharpe and Strangway, 1976; Solomon, 1977) or deep-mantle heat sources (Cassen *et al.*, 1976) has been suggested. Late core formation would support a cool initial state for Mercury but is at variance with accretion models (e.g., Schubert *et al.*, 1988). The most likely reason for Mercury not having a totally frozen core, however, is the incorporation of a light alloying element into its core that reduces the core melting temperature. Here, sulfur is the most likely candidate (Ringwood, 1977; McCammon *et al.*, 1983). The amount of sulfur depends on the formation scenario of Mercury. In the most conservative equilibrium condensation models there is no sulfur at all in the Mercurian core due to the close position of the planet to the Sun (Lewis, 1972; Grossman 1972), an assumption mostly used in earlier thermal evolution models (Siegfried and Solomon, 1974; Sharpe and Strangway, 1976, Solomon, 1976, 1977). However, a low concentration of sulfur has been suggested with radial mixing of planetesimals and/or nonequilibrium condensation models (Basaltic Volcanism Study Project, 1981; Wetherill, 1985). Mercury's inner structure that is consistent with its high density can actually be explained with a wide range of the sulfur concentration in the core ranging from a pure iron core to a core with eutectic Fe–FeS composition (Harder and Schubert, 2001).

The observation of the magnetic field provides probably the best constraint on the interior evolution and assuming that it is indeed generated in the core and not a remanent field – as on Mars – it proves that there is a present convecting outer core. The existence of the fluid outer core places constraints on the maximal sulfur content in the core. The inferred small planetary contraction from the lobate scarp, on the other hand, places constraints on the minimal amount of sulfur in the core (Hauck *et al.*, 2004). In general, one can state: the less efficient the planetary cooling the smaller is the required sulfur content.

Parametrized convection models by Schubert *et al.* (1988) that base on constant-viscosity laws suggest a concentration between 2 and 7 wt.% of sulfur to have a present-day outer fluid core and to satisfy further the geological constraint of small planetary contraction over time. If the sulfur content is >7 wt.%, no inner core forms to help drive convection in the outer core and the constraint of a small planetary contraction suggests a bulk core sulfur content larger than 2 wt.%. The models further assume a dry mantle and a depletion of potassium as suggested by geochemical models (Basaltic Volcanism Study Project, 1981). Two-dimensional (2-D) and 3-D mantle convection models with strongly temperature- and pressure-dependent rheology and a heat-source density consistent to the silicate vaporization model (Fegley and Cameron, 1987) by Conzelmann (1999) and Buske (2006) show slower mantle cooling in comparison to the models based on constant-viscosity laws (Schubert *et al.*, 1988). As a consequence, they predict a concentration of *c.* 1–3 wt.% S with a relative inner-core size smaller than 0.6 (**Figure 5**). The models suggest further sluggish convection in the deep mantle throughout the entire evolution (**Figure 6**).

Recent thermochemical evolution models based on temperature-dependent viscosity laws that couple the thermal, magmatic, and tectonic evolution of Mercury (Hauck *et al.*, 2004) indicate that a dry non-Newtonian mantle allows the planet to cool more efficiently than a dry Newtonian mantle. As a consequence, acceptable models for the sulfur content range between 6.5 and 7.5 wt.% S to explain the observations. Their models also suggest an initially convecting mantle only during the early stages of evolution accompanied by extensive melting and differentiation. Subsequently, convection and melting have been absent and the planet has been cooling down in a conductive regime. The onset time to a conductive regime depends also on the efficiency of mantle cooling due to crust formation. The more the crust is produced, the stronger is the mantle cooling and the associated depletion of the mantle in heat sources. This on the other hand implies an earlier change – or a change at all – to a conductive regime and a higher sulfur content in the core.

A conclusion drawn from these studies (Schubert *et al.*, 1988; Conzelmann, 1999; Hauck *et al.*, 2004; Buske 2006; Breuer *et al.*, 2007) is that the present state of the mantle, that is, convective or conductive, the sulfur content in the core and the inner-core size cannot be constrained by the observations. Too

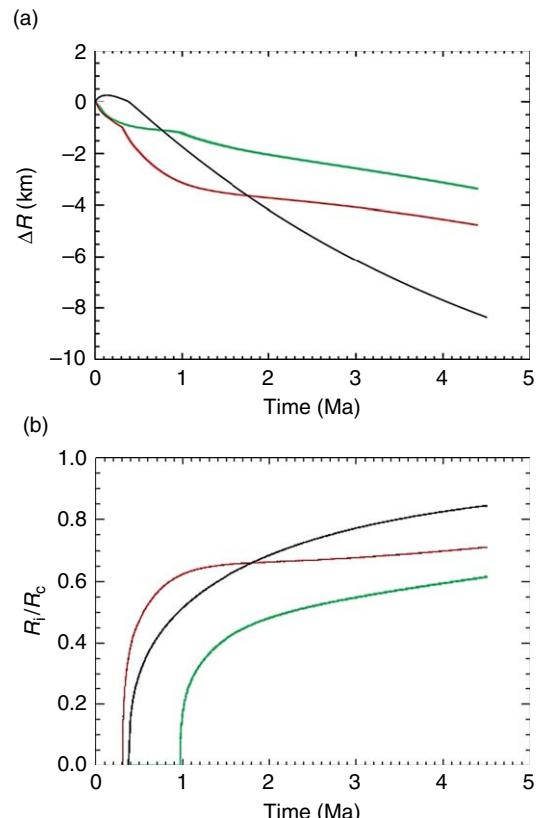


Figure 5 Contraction of the Hermian surface (a) and the growth of the inner core (R_i/R_c) (b) as a function of time for a parametrized convection model based on constant-viscosity law (black line), for a 2-D axisymmetric convection model with strongly temperature-dependent viscosity (red line) and a 2-D axisymmetric convection model with strongly temperature- and pressure-dependent viscosity (green line). The viscosity increases by a factor of 10 in the convecting mantle. Models assume a concentration of radioactive heat sources consistent to the silicate vaporization model (Fegley and Cameron, 1987) and 1% of sulfur in the core. Modified from Conzelmann (1999) Thermische Evolution des Planeten Merkur berechnet unter Anwendung verschiedener Viskositätsgesetze. PhD Thesis, University of Münster.

many parameters, which influence significantly the thermal evolution, are unknown such as the mantle rheology and the amount and distribution of the radioactive heat sources. However, the studies suggest a dry mantle and a strong depletion of potassium in the mantle to satisfy both the geological observations of small contraction and a present fluid outer core.

The thermal evolution models (Schubert *et al.*, 1988; Hauck *et al.*, 2004; Conzelmann, 1999; Buske, 2006) infer the following evolution of the magnetic

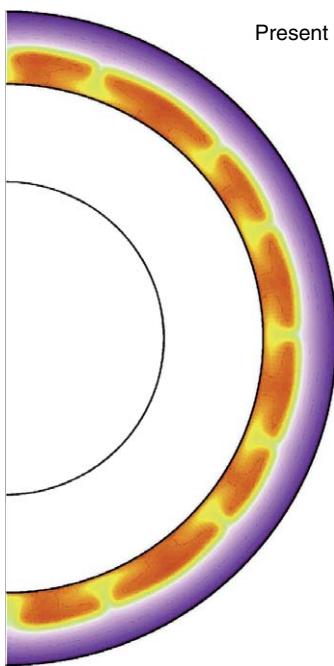


Figure 6 A typical temperature field of the convecting mantle (colored region) in Mercury calculated with a 2-D axisymmetric convection model with strongly temperature- and pressure-dependent viscosity. The temperature decreases from yellow and violet. The upper violet part is the stagnant lid, convection takes place below. The solid circle in the center indicates the present size of the inner core. The model is the same as in [Figure 5](#) with green line. Modified from [Conzelmann \(1999\)](#) Thermische Evolution des Planeten Merkur berechnet unter Anwendung verschiedener Viskositätsgesetze. PhD Thesis, University of Münster.

field: in the early evolution, a dynamo is generated by thermal convection before the start of the inner core and a chemical dynamo. Whether the dynamo action ceases for a time until inner-core growth provides sufficient energy depends on the onset time of inner-core growth. The higher the sulfur content the later is the onset of inner-core growth (e.g., [Schubert et al., 1988](#); [Hauck et al., 2004](#)). A detection of remanent magnetized crust and the determination of its age by future missions could therefore help to constrain the amount of sulfur in the core and thus the formation scenario.

The hydromagnetic dynamo as assumed in the thermal evolution models above poses a problem in the strength of the calculated field. Although there is presently no satisfying parametrization to calculate the magnetic field strength, estimates suggest a much larger field than that observed ([Stevenson et al., 1983](#);

[Schubert et al., 1988](#)). Recent thin-shell dynamo models, however, have shown that planets with a large solid inner core (relative inner core size larger than about 0.8) can produce magnetic fields with Mercury-like field intensities ([Heimpel et al., 2005](#); [Stanley et al., 2005](#)). Alternatively, a deep dynamo model that suggests a rather small inner core (smaller than 1000 km in radius) has been proposed by [Christensen \(2006\)](#). In this model, the dynamo operates only at depth and the associated dynamo field is strongly attenuated by the skin effect through a stable conducting region of the upper core. Another alternative way of generating a weak magnetic field in Mercury is by a thermoelectric dynamo ([Stevenson, 1987](#); [Giampieri and Balogh, 2002](#)). This dynamo makes use of a thermoelectric electromotive force set up at a distorted CMB. Such a dynamo requires topography variations of the CMB of the order of 1 km due to mantle convection. Whether this kind of dynamo is active in Mercury cannot be concluded from the current magnetic data. It is, however, possible to detect core–mantle undulations by inversion of the global Mercurian gravity and topography field ([Spohn et al., 2001](#); [Breuer et al., 2007](#)). New insights into the magnetic field evolution and the dynamo mechanism of Mercury are expected from the ongoing MESSENGER mission.

Detailed geologic mapping of the observed surface has shown that large areas of relatively ancient inter-crater plains may indicate that more extensive volcanism accompanied the heavy-bombardment period on Mercury than on the Moon. Its smooth plains are nearly as reflective as its heavily cratered regions, which has led to controversy over the origin (volcanic or otherwise?) of the smooth regions. Estimates of the crustal thickness have a large possible range. [Anderson et al. \(1996\)](#) used the observed center-of-mass center-of-figure offset together with an assumption of Airy isostasy to infer a crustal thickness of 100–300 km. Further constraints on the crustal thickness have been derived from viscous topographic relaxation ([Nimmo, 2002](#)). If the faulting observations provide a reasonable guide to heat fluxes on Mercury at 4 Ga, topography is most likely to have survived if the mean crustal thickness were less than 200 km. Studies employing topographical information about the Mercurian surface ([Nimmo and Watters, 2004](#); [Watters et al., 2004](#)) suggest that the extensional faulting of observed lobate scarps is consistent with a mean crustal thickness of 90–140 km.

Thermochemical evolution models with parameterized stagnant lid convection and assuming an

Earth-like mantle for Mercury do not produce crust as voluminous as the largest of the above values (Hauck *et al.*, 2004; Breuer *et al.*, 2007). The derived crustal thicknesses vary between 5 and 70 km depending on the initial temperature distribution, the assumed amount of radioactive elements, and the mantle rheology. If one neglects the uncertainties in such models, either a peridotite-dominant mantle is inappropriate or much of the crust is primordial. A primordial crust is supported by the reflectance spectrum of the surface of Mercury. The spectrum is similar to that of the lunar highlands (Vilas, 1988; Sprague *et al.*, 1997), which are predominately plagioclase and are most likely formed by freezing of a magma ocean. Furthermore, the radar characteristics of the surface are also reminiscent of the lunar highlands (Harmon, 1997).

Recent thermochemical evolution models (Breuer *et al.*, 2007) that include the influence of a low-conductivity crust have shown that a global partial melt zone underneath the stagnant lid of about 200 km may have been present during a long period of time possibly extending to the present day. It is, however, questionable whether and how much of this melt could rise toward the surface. If there is a global layer of partially molten mantle, tidal dissipation in this region of Mercury may provide an important heat source. The influence of a low-conductivity crust and the potential energy supply by tidal dissipation are in accord with the observation of small contraction since the end of heavy bombardment and a convecting deep mantle is more likely.

10.09.6.2 Venus

The statistically uniform distribution of impact craters prevents us from identifying older or younger regions and thus from determining resurfacing rates. The surface has a global average age of 300–800 Ma (Schaber *et al.*, 1992; McKinnon *et al.*, 1997). From geochemical measurements of the Soviet landers, the surface appears to be basaltic (McKenzie *et al.*, 1992a); an observation that is supported by Magellan radar images. These images show many volcanic flows that are similar to basaltic flows on Earth. It has been suggested (McKinnon *et al.*, 1997) that the entire surface was covered by a volcanic event of global scale, but such an event is not required by the crater distribution. In such models, the resurfacing event was followed by almost no volcanic activity to the present day. The cause for this global resurfacing event and whether it was a single event or

one of a periodically recurring set of events is highly uncertain. Thus, a number of papers that speculate on the global resurfacing event can be found in the literature (see references in Bougner *et al.* (1997) and Nimmo and McKenzie (1998)).

1. Widespread melting is caused by a transition from layered to whole-mantle convection (Steinbach and Yuen, 1992). As Venus is cooling the tendency for mantle layering that is caused by the endothermic spinel-to-perovskite phase transition decreases until the layered structure breaks up with a single catastrophic event.
2. Periodic instabilities of a depleted layer beneath the crust (Parmentier and Hess, 1992; Herrick and Parmentier, 1994) result in the global melting event. In that scenario, the competition between compositional stratification due to crust formation and thermal buoyancy may cause episodic overturn of the mantle.
3. The resurfacing is associated with a transition from oscillatory to steady convection due to planetary cooling and thus a decreasing Rayleigh number (Arkani-Hamed *et al.*, 1993).
4. The global resurfacing has been related to some kind of episodic plate tectonics. After a resurfacing event, the cold lithosphere has been subducted, and a new lid grows conductively. As this lid thickens, its negative buoyancy increases, until at some critical thickness, it is able to subduct again, and thereby instigates a new resurfacing event (Turcotte, 1993; Weinstein, 1996; Fowler and O'Brien, 1996; Moresi and Solomatov, 1998). Herrick (1994) concluded that a change in surface boundary condition could lead to a rapid rise in mantle temperature and catastrophic resurfacing in form of plate subduction.

5. Solomatov and Moresi (1996) assume that the resurfacing event is caused by the cessation of plate tectonics. The reduction in convective stresses with time might cause deformation in the stagnant lid to change from brittle (plate tectonics) to ductile (little surface movement).

The commonly assumed rapid global resurfacing event has been questioned by Campbell *et al.* (1999) and most recently by Bond and Warner (2006). According to the latter authors, the cratering record allows a variety of interpretations in terms of volcanic resurfacing including a global decrease with time in the rate of an otherwise statistically distributed volcanic activity. Previous authors had concluded that Venus underwent a major transition in tectonic style albeit more gradual than the previously

postulated sudden global resurfacing event. The surface geology also seems to indicate that modifications of the surface are planet-wide and gradual over long periods of time rather than episodic and locally confined (Ivanov and Head, 2006). Whether there is present-day volcanic activity on Venus is unclear. Some sulfuric clouds have been interpreted as being evidence of recent volcanism (Fegley and Prinn, 1989). Recent data from the infrared spectrometer VIRTIS on Venus Express (Helbert *et al.*, 2006) suggest differences in surface temperature that can be linked to morphologic features on the surface, but it is too early for more far-reaching conclusions. It is, however, assumed that Venus is an active planet in its interior. The significant correlation between the geoid and topography at low degrees, together with a relatively high admittance ratio, may indicate that a significant portion of the long-wavelength topography has a dynamic origin by mantle convection (Simons *et al.*, 1997; Pauer *et al.*, 2006). However, it should be noted, as numerical investigation of convection in strongly temperature-dependent viscosity fluids has progressed, it has become apparent that temperature variations in the nearly stagnant conducting lid are the primary source of topography (Solomatov and Moresi, 1996), and not dynamic stresses, which dominate at low Rayleigh number and small viscosity contrast. Thus topographic rises are not so much held up by upwellings as they are buoyed by the heat delivered to the lithosphere. Plume activity has been suggested by Moore and Schubert (1997), using a model of thermal isostasy. The authors found a very thick thermal lithosphere (200–400 km) which is locally thinned by up to a factor of 5 beneath the volcanic rises. This was attributed to convection in Venus' mantle at a Rayleigh number of about 10^7 and a large viscosity contrast based on the dynamical calculations of Solomatov and Moresi (1996). Similar conclusions are given by Phillips (1994) and Kucinskas and Turcotte (1994).

The influence that the resurfacing event must have had on the thermal history of Venus depends actually on the resurfacing mechanism. In comparison to the Earth, the question of early plate tectonics on Venus, which has been suggested by several papers, is particularly interesting. But unlike for Mars, there is little present evidence for it. However it is likely that the main topographic features of plate tectonics, that is, ridges, trenches, and transform faults, may disappear during the 500 Ma after plate tectonics has stopped. Plate tectonics on Earth can take place because the driving forces (i.e.,

slab pull and ridge push) exceed the resistive forces (i.e., fault friction, flexure, and viscous drag) (Forsyth and Uyeda, 1975). It is commonly suggested that water on the surface of Earth reduces the resistive forces by modifying the rheology of the cold plates such that subduction becomes feasible. The high D/H ratio in the Venusian atmosphere suggests that Venus had substantial water in the past (Fegley, 2004). Today Venus is desiccated as far as the available spacecraft data can tell. Thus, present plate tectonics is not feasible if water on the surface is required (Nimmo and McKenzie, 1998). Another possible explanation for the present-day lack of plate tectonics on Venus is that the high surface temperature and thick basaltic crust make the lithosphere less dense than the underlying convecting mantle – subduction is inhibited (Herrick, 1994). Furthermore, a high mantle viscosity might cause a large viscous drag force on any moving plates (McKenzie, 1977). On Earth, this mantle drag is greatly reduced by the effect of the low-viscosity zone. For Venus, it has been suggested on the basis of the interpretation of gravity and topography data, that a low-viscosity zone beneath the lithosphere does not exist (Kiefer *et al.*, 1986; Nimmo and McKenzie, 1996). Recent models by Pauer *et al.* (2006) show that the best fit to the data imply a gradual increase in viscosity.

Another important aspect of the thermal evolution of Venus is the coupling of the atmosphere or hydrosphere with the dynamics of the solid interior. The thick CO₂ atmosphere of Venus results in enhanced surface temperatures due to the greenhouse effect. The surface temperature which is the upper thermal boundary condition of the convecting interior, on the other hand, influences the heat transport and thus the thermal evolution. These effects have been neglected so far in most evolution models. A first attempt has been made by Phillips *et al.* (2001) who investigated a coupled evolution for Venus using a simple parameterized mantle convection model (including partial melting) and a radiative–convective atmospheric model. Feedback was incorporated by release of water associated with melt extraction into the Venusian atmosphere which adds to the greenhouse effect and thus affects the surface temperature. After 2 billion years of evolution, the coupled model was hotter by several tens of kelvins and its extrusive volcanic flux was 4 times higher than the case without coupling. Their study showed a complex interplay between Venus convective evolution, volcanic activity, and atmospheric state even with such a simple model featuring only basic processes. More

complex and elaborate models are required to obtain a more detailed understanding of the interactions.

Volcanic features on Venus include several large highlands topped by massive volcanoes, a large number of comparatively small shield volcanoes that are apparently randomly distributed, volcanic plains, and about 500 coronas (circular large-scale volcanotectonic features). It is widely accepted that coronas form above upwelling plumes (Smrekar and Stofan, 1999). Johnson and Richards (2003) have recently reanalyzed the corona distribution and have concluded that corona activity is concentrated in the Beta, Atla, and Themis rift systems.

The thickness of the crust has been constrained using gravity and topography data and estimates of the average thickness are about 35 km (Simons *et al.*, 1994; Konopliv and Sjogren, 1994). Thermal evolution models with parametrized convection suggest that Venus may have a 100-km-thick lithosphere which may consist largely of basalt (Spohn, 1991). Those models, however, assume one-plate tectonics throughout the entire evolution; that is, crust was not recycled either due to plate tectonics or some other mechanism at the time of resurfacing. Furthermore, as Turcotte (1989) has noted, the thickness of the Venusian crust may be limited by the basalt-to-eclogite phase transformation which occurs at depths of 30–70 km, depending on the temperature (Ito and Kennedy, 1971; Ringwood, 1975). The dense eclogite phase is gravitationally unstable with respect to the mantle below. When the crust is sufficiently thick, crustal delamination may occur. Crustal thickness variations exist between the various geological provinces: estimates are between 20 and 50 km at the volcanic rises (Phillips, 1994; Smrekar, 1994), 20 and 40 km in the plains (Konopliv and Sjogren, 1994; Grimm, 1994), and 35 and 60 km around tessera sites (Kucinskas and Turcotte, 1994; Simons *et al.*, 1994; Grimm, 1994). The elastic thickness of the lithosphere ranges between 10 and 50 km (Sandwell and Schubert 1992; Johnson and Sandwell, 1994; Brown and Grimm, 1996; McKenzie, 1994; Smrekar, 1994; Phillips, 1994; McKenzie and Nimmo, 1997), with a clustering around 30 km. Anderson and Smrekar (2006) demonstrate that models which include bottom loading show a larger range of local scales in the crustal thickness and the elastic lithosphere thickness than previous models. The results indicate that geologic processes may be very complex on Venus.

Since the first flyby of Venus by Mariner 2 in 1962 at a distance of 6.6 planetary radii, it was clear that

Venus does not have an Earth-like magnetosphere. In fact, the following mission of Mariner 5, in 1967, placed an upper limit on the magnetic dipole moment of Venus of $\sim 10^{-3}$ that of Earth's. Later, most definitive measurements of the magnetic moment of Venus were obtained during the Pioneer Venus Orbiter mission in its first years of operation (1979–81). The upper limit on the dipole moment obtained from the Pioneer Venus Orbiter placed the Venus intrinsic magnetic field at less than about 10^{-5} times that of Earth (Russell *et al.*, 1979a, 1979b).

The lack of a present dynamo does not imply that Venus never had an intrinsic magnetic field although we have no information that relates directly to the past history of the field. The surface temperature of about 740 K is respectively close to the Curie temperature (T_{curie}) of the main magnetic carriers on terrestrial planets, such as magnetite, hematite, and pyrrhotite with T_{curie} of 850, 940, and 600 K, respectively (e.g., Dunlop and Özdemir, 1997). Thus, temperatures of the crust, except possibly in an upper thin layer, are expected to be higher than the Curie point (at temperatures below the Curie-point magnetization could persist in rocky materials). Any remanent crustal magnetic field from an early period of dynamo activity might be weak on Venus if existent at all. It is also important to note that the slow rotation of Venus (a Venus day of ~ 243 Earth days is almost equal to the length of its year of ~ 224 days, and its sense of rotation is retrograde) does not exclude dynamo action.

Thermal evolution models suggest that there was a magnetic moment of Venus of the same order as Earth's for about the first 3 billion years of Venus' history (Stevenson *et al.*, 1983; see Figure 7). During that time, thermal convection drove the dynamo as suggested for the other terrestrial planets. After the core cooled below the adiabatic heat flow of the core, the magnetic field ceased and did not rejuvenate; the core was too hot and the pressure is too low for inner-core growth. The evolution models by Stevenson *et al.* (1983) have been calculated with the parametrization based on constant-viscosity laws which are more representative for planets with lithosphere delamination (see Sections 10.09.4.1 and 10.09.4.2). Assuming that Venus was in a stagnant lid regime throughout the entire evolution, the phase of early dynamo action would be significantly reduced because core cooling might be overestimated with the former models.

An alternative scenario for the present lack of a Venus magnetic field suggests that Venus's field

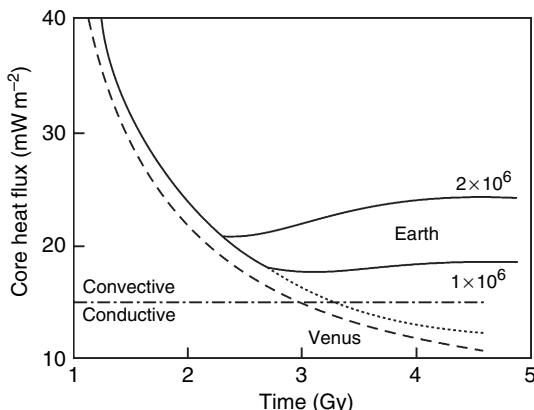


Figure 7 The heat flow out of the cores of Earth (solid and dashed lines) and Venus (dotted line) as a function of time. The curve parameter for the models is the specific energy release upon inner-core freezing in J kg^{-1} . The dash-dot line marks the conducted heat flux along the core adiabat. Without inner-core solidification, larger heat fluxes indicate thermal convection in the core. Smaller heat fluxes are then indicative of cessation of thermal convection and dynamo generation. With inner-core growth, chemical buoyancy helps to drive convection and the core heat flux may be sub- or superadiabatic. Modified from Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.

ceased as Venus transitioned from a plate tectonic to a stagnant lid regime around 700 Ma ago (Stevenson, 2002). As suggested for Mars (Nimmo and Stevenson, 2000), the transition heated the mantle and the core. If Venus had an inner core but a mantle temperature increasing in time, then the inner core would have ceased to grow (it may actually be shrinking in radius) and the dynamo could not be driven by chemical buoyancy released upon inner-core growth.

10.09.6.3 Moon

The thermal evolution of the Moon has been a controversial subject. Early models start the evolution of the Moon with an accretionary initial temperature profile. This temperature profile increases from a relatively cold deep interior (less than 1000 K) to reach a maximum in the upper mantle from where temperature decreases toward the surface (Matsui and Abe, 1986). Similar temperature profiles have been used in other models of the Moon's thermal evolution (Cassen and Reynolds, 1973; Toksöz *et al.*, 1978; Cassen *et al.*, 1979). Some of these authors have assumed a magma ocean overlying a cold interior. There is evidence, however, from the trace-element pattern and the remanent magnetization of the

Moon's surface rock that the Moon started hot or that the interior heated rapidly after accretion (Runcorn, 1977; Palme *et al.*, 1984; Binder, 1986). The mechanism by which the lunar interior may have been rapidly heated has been the subject of debate. Among possible mechanisms that have been discussed are the decay of short-lived radionuclides such as ^{26}Al , tidal heating, and electromagnetic heating, none of which are very satisfying. The hypothesis of a giant impact origin of the Moon (see Chapter 10.14), however, provides a way of accreting the Moon hot (Stevenson, 1987) if the Moon forms more rapidly than the vapor that is ejected from the Earth cools.

More recent evolution models assume an initially hot interior but the questions about the thermal and structural state of the Moon after freezing of the magma ocean and whether convection is an important form of heat transport in the lunar interior are still open. It is generally accepted that the anorthositic crust of the Moon formed as a floating crust on a magma ocean. Estimates of the depth of the primordial magmasphere, however, range from the whole-Moon melting to thin melt layers above partially molten zones. Assuming that the magma ocean freezes rapidly during *c.* 100–200 Ma (Solomon and Longhi, 1977; Minear and Fletcher, 1978) without any disturbances, this part of the mantle most likely became chemically stratified following magma ocean differentiation and crystallization with the late dense iron-rich phases coming to rest upon a less-dense Mg-rich phase. Furthermore, most of the incompatible and heat-producing elements would have been concentrated in a KREEP-rich layer that is located just below the anorthositic crust, leaving the mantle strongly depleted in radioactive heat sources (Figure 8). Such a layering has strong implications for the subsequent evolution of the Moon since it is prone to convective overturn and mixing (Hess and Parmentier, 1995; Alley and Parmentier, 1998). It has also been speculated that during most of the Moon's history (i.e., after the first rapid overturn) heat might be transported by conduction alone (Kirk and Stevenson, 1989; Pritchard and Stevenson, 1999; Wieczorek and Phillips, 2000) as a consequence of the strong depletion of radioactive elements and the stable chemical layering in the mantle.

Whether the rapid convective overturn scenario took place and whether a subsequent stable chemical layering developed depends on the initial depth of the magma ocean and the freezing time of the magma ocean. Spohn *et al.* (2001) show with 2-D and 3-D

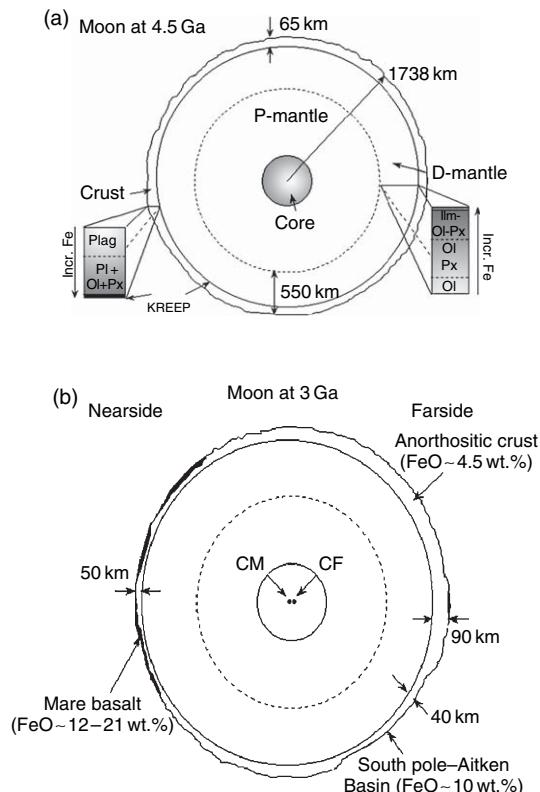


Figure 8 (a) A schematic cross-section (roughly to scale) through the Moon immediately after crystallization of the magmasphere. The dashed line represents the 550 km seismic discontinuity. P-mantle is primitive mantle (unmelted or partially melted). D-mantle represents differentiated mantle. The sequence of cumulates in the upper mantle shows an upward enrichment in iron with a KREEP layer at the top sandwiched between the crust and mantle. The crustal cumulates are presumed to become more mafic and more iron-rich with increasing depth. (b) A schematic cross-section through the Moon at 3 Ga. Note the thick farside crust and the giant South Pole Aitken basin. CM, center of mass; CF, center of figure. The thickness of mare basalts on the nearside is exaggerated. The FeO concentrations are based on Clementine data (Lucey *et al.*, 2000). Modified from McCallum (2001) A new view of the Moon in light of data from Clementine and Prospector missions. *Earth, Moon and Planets* 85–86: 253–269. With kind permission from Springer Science and Business Media.

convection models that the deep layers of a 700-km-deep magma ocean were kept (partially) molten over a significant part of lunar history. It is not well understood whether an unstable chemical layering would develop under these conditions. Furthermore, it is likely that the deep layers became chemically rejuvenated by the convective mass flow. The situation is different for a shallow magma ocean. In the case of a 300-km-deep magma ocean, a melt zone forms below

the magma ocean's original depth in the previously solid residual and convecting mantle; the layering of the freezing magma ocean would most likely not be disturbed but it is uncertain whether the unstable dense layers are thick enough to initiate a rapid overturn. In the case of thermal convection of the lunar interior, the evolution is found to be characterized by the growth of a massive 700–800-km-thick lithosphere while the lower mantle and core cool only by a few hundred kelvins (Spohn *et al.*, 2001); this is the typical evolution of a one-plate planet.

The Moon has been divided into two major geological terranes: (1) the highlands and (2) the maria. The highlands are more cratered and higher in elevation than the darker maria. The bulk of the highland crust rocks, consisting of ferrarite anorthosites, is formed as a floating crust on a magma ocean during the first 100–200 Ma (Lee *et al.*, 1997). The maria are impact craters filled by basaltic lava. The basaltic lava, the most prominent witnesses of volcanic activity on the Moon, flooded some (but not all) of the major impact basins some 100 Ma after impact basin formation (Figure 8). The ages of large mare basalts dated radiometrically range from 3.9 to 3.2 Ga. However, the onset of basaltic volcanism appears to be earlier; an age of 4.2 Ga has been reported for a basalt clast in a highland breccia (Taylor *et al.*, 1983). Crater-counting methods indicate basalts in some unsampled areas of the Procellarum basin are as young as 1.3 Ga (Hiesinger *et al.*, 2000).

There is geochemical evidence (Hess, 2000) of a deep origin for the mare basalts having relatively primitive compositions. This has often been cited as evidence for deep remelting of the lunar mantle. Furthermore, the TiO₂ content decreases with increasing depth and with increasing age (Basaltic Volcanism Study Project, 1981, Head and Hiesinger, 1999), suggesting that the mare basalt source region moves deeper with increasing time. This observation has been confirmed with thermal evolution models using 2-D and 3-D mantle convection (Spohn *et al.*, 2001). The results show a melt zone freezing from above as heat is removed through the growing lithosphere (Figure 9). The freezing of the melt zone from above suggests that the magma source region moves to greater depths with time.

The time delay between mare formation and mare fill of about 100 Ma is not well understood. This time delay may be due to the time necessary for the melt to reach the surface after the impact since some time is required for a magma diapir or dike to penetrate through the lithosphere. Furthermore, ascending

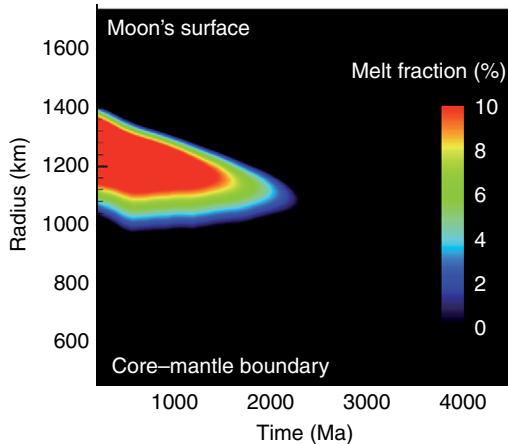


Figure 9 Partial melt zone in the Moon's mantle as a function of time. Modified from Spohn T, Acuña MA, Breuer D, Golombek M, et al. (2001a) Geophysical constraints on the evolution of Mars. *Space Science Reviews* 96: 231–262.

mantle diapirs likely stall at a basal crustal neutral buoyancy zone (Head and Hiesinger, 1999) before reservoir overpressurization propagates dikes to the surface. Alternatively, it has been suggested that the time delay is the consequence of a chemically stratified mantle following magma ocean differentiation and crystallization (Hess and Parmentier, 1995; Alley and Parmentier, 1998). In the presence of a stabilizing compositional density gradient, the onset of thermal convection (due to heating of the interior, the thermal buoyancy can increase with time and overcome the stabilizing compositional density gradient) and thus the volcanic activity will be delayed if the model is started from a nonconvecting state (Alley and Parmentier, 1998). However, if it is assumed that the onset of convection was delayed until tens of million years after the formation of the maria, the observed volcanic activity prior to the period of mare fill will be left unexplained.

Another open issue of the volcanic activity on the Moon is the observation that the mare basalts are mainly located on the nearside, the side facing the Earth. Thus, like Mars, the Moon shows a hemispherical asymmetry that is also reflected in a variation in the crustal thickness: the nearside with most of the mare basalts shows a significantly thinner crust in comparison to the crust on the farside. This was first referred from the observed 2 km offset between the Moon's center of mass and figure roughly along the Earth–Moon axis (Kaula et al., 1972; see Figure 8(b)). The nearside–farside dichotomy of crustal thickness has been confirmed from

Clementine data (Nozette et al., 1994). One explanation for the existence of mare basalts in the thinner crust of the nearside is that the hydrostatic pressure of the mare source controls the eruption of mare basalts. Assuming that the maximum depth of the mare source was globally uniform, then mare basalts would only be able to erupt at the surface below a critical elevation. Thus, the higher elevations of the lunar farside could have prevented farside magmas from reaching the surface due to their lack of the necessary pressure in their source. This scenario might, however, be a bit simplistic since the South Pole Aitken Basin has the lowest elevations on the Moon but mare flows in this basin are volumetrically insignificant when compared to the nearside basins.

It has been suggested that the basic lunar asymmetry was established during the magma-ocean stage (Warren and Wasson, 1979; Loper and Werner, 2000) since the dynamically unstable configuration of the fractionated magma ocean might result in a large-scale overturn event (Hess and Parmentier, 1995). The large-scale downdwelling eventually focused on the nearside would recycle the evolved KREEP-rich material into the deeper mantle providing a heat source for subsequent internal remelting on the nearside (Parmentier et al., 2002; see Figure 10). However, the timing, lateral extent, and even existence of the postulated overturn event are not well constrained.

Based on recent γ -ray data from the Lunar Prospector it has been realized that the crustal dichotomy is in fact more restricted (Lawrence et al., 1998). About 60% of the Moon's mare basalts reside in the Procellarum and Imbrium regions (Wieczorek and Phillips, 2000). This province is in addition a unique geochemical crustal province enriched in incompatible and heat-producing elements. Wieczorek and Phillips (2000) have argued for a genetic relationship between the two. Specifically, by distributing a layer of KREEP basalt inhomogeneously in the lunar interior, their thermal models predict that the lunar mantle should have partially melted only beneath this province. The eruption of mare basalts thus may be primarily controlled by the distribution of heat sources in the Moon.

Paleomagnetic data, combined with radiometric ages of Apollo samples, suggest that a field of possibly 10^5 nT existed at 4.0 Ga decreasing to 5×10^3 nT at 3.2 Ga (Stephenson et al., 1975; Cisowski and Fuller, 1986). Because the present magnetic field of the Moon is negligible, Runcorn (1975) has argued that

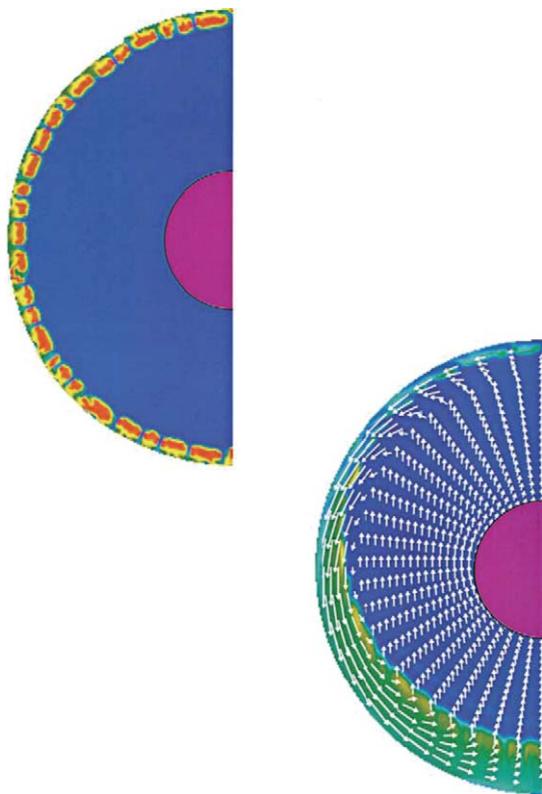


Figure 10 Finite-amplitude Rayleigh–Taylor instability of an initially 100-km-thick layer overlying an interior region in which the viscosity increases by a factor of 5000 over a depth interval of 400 km. Early time is on the left; later time is on the right. Velocity vectors show the spherical harmonic degree-1 flow pattern present at the later time. Central region corresponds to a dense liquid metallic core. Consistent with expectations from linearized stability analysis instability begins at a wavelength comparable to the layer thickness; but, at a much longer wavelength spherical harmonic degree-1 flow subsequently grows more rapidly because of its longer wavelength. The early short-wavelength instability creates an effectively thicker dense layer that is then unstable at long wavelength. Modified from Parmentier EM, Zhong S, and Zuber M (2002) Gravitational differentiation due to initial chemical stratification: Origin of lunar asymmetry by the creep of dense KREEP? *Earth and Planetary Science Letters* 201: 473–480.

the lunar rocks were magnetized at the time of their origin by a field of internal origin. The easiest explanation for such a field is the operation of a dynamo in an iron-rich lunar core. Some researchers doubt that an internal dynamo is required to explain the magnetization of the Moon, and favor an alternative idea; the observed magnetic signature is suggested to be generated in association with large impacts during early history (Hide, 1972; Hood and Vickery, 1984; Hood and Huang, 1991; Hood *et al.*, 2001). In this

model, magnetization should be concentrated diametrically opposite to major impact basins. Satellite observations from the Lunar Prospector mission show concentrations of crustal magnetization at the antipodes of some but not all of the large impact structures.

Assuming an internal origin of the lunar magnetic field, one must be aware that the paleomagnetic data show a gap of about 500 Ma in the remanent magnetization of the lunar rock. There exist two alternative explanations for this gap: (1) the remanent magnetization of older basaltic rock was destroyed by impact gardening and/or the volcanic activity during the first 500 Ma was very low or (2) the internal field started about 500 Ma after core formation.

An early magnetic field has been suggested from thermal evolution models incorporating 2-D and 3-D convection with strongly temperature-dependent rheology (Konrad and Spohn, 1997; Spohn *et al.*, 2001). These models show that a thermally driven dynamo might have been active from the beginning of the core formation up to about 3 Ga if the core was superheated by about 200 K with respect to the mantle. These thermal evolution models, however, assume that at least the lower mantle is not chemically stratified.

A late onset of dynamo action takes into account the thermal and chemical consequences of the magma ocean (Stegman *et al.*, 2003). After crystallization of the magma ocean, a dense ilmenite and pyroxene cumulate located just underneath the enriched KREEP layer sinks to the deep lunar interior. It has been suggested that part of the KREEP layer is carried away together with the dense layer. This mixed enriched layer encircles the lunar core and insulates it from the rest of the mantle, trapping heat in the core and preventing the core from cooling convectively, and also from developing a dynamo. After a period of time, the radiogenic material within this ‘thermal blanket’ decays and heats the lower layer. Eventually the material becomes more buoyant than its surroundings and rises toward the surface. With the removal of the thermal blanket, the core is then able to convect vigorously to cool itself, and this can produce a short-lived dynamo. An essential element of the model is the timing. The length of time that it takes the thermal blanket to heat up and rise back toward the surface is broadly consistent with two important events in early lunar history: the eruption of the mare basalts onto the lunar surface and the magnetization of lunar rocks. Even so, the model has some shortcomings. For this blanketing

layer to become thermally unstable so that core convection can be initiated the density difference between the lower layer and the overlying mantle can only be a few tens of kilograms per cubic meter. A much higher density difference, however, is suggested from models of Rayleigh–Taylor instability for that layer to sink fast enough into the deep interior (Parmentier *et al.*, 2002).

10.09.6.4 Mars

A fundamental problem in the evolution of Mars is the timing and the origin of the crustal dichotomy. The southern highlands and northern lowlands of Mars differ markedly in average elevation (Smith *et al.*, 1999) and crustal thickness (Zuber *et al.*, 2000; Neumann *et al.*, 2004). Although it is generally accepted that this crustal dichotomy is one of oldest features on Mars, the exact timing of the dichotomy formation, which has implications for the formation mechanism, is strongly debated. Early studies on faulting and other tectonic features at the crustal dichotomy boundary suggested a Late Noachian/Early Hesperian (3.7 Ga) formation for the dichotomy (McGill and Dimitriou, 1990). Most recent studies suggest an Early Noachian or earlier (>3.9 Ga) formation time for the dichotomy (Frey *et al.*, 2002; Nimmo and Tanaka, 2005). One argument for this early crust formation is the finding of quasi-circular depressions (QCDs) in the northern lowlands (Frey *et al.*, 2002). They are interpreted as buried impact craters and are suggestive that the basement of northern lowlands formed during the Early Noachian and is even older than the visible highlands. Solomon *et al.* (2005) even suggest an early crustal formation, including the dichotomy structure, as early as the first 50 Ma after the solar system formation. His lines of evidence include the isotopic anomalies in Martian meteorites (Chen and Wasserburg, 1986; Harper *et al.*, 1995; Borg *et al.*, 1997; Brandon *et al.*, 2000; Halliday *et al.*, 2001) that indicate an early differentiation event of the mantle about 4.5 Ga. The SNC meteorite isotope data further suggest insignificant reservoir mixing thereafter.

The origin of the crustal dichotomy has variously been related to external (Wilhelms and Squyres, 1984; Frey and Schultz, 1988) and internal processes (Wise *et al.*, 1979), but none of the proposed formation mechanisms has been fully convincing in part due to the uncertainty in the timing of the dichotomy formation. For the external processes, one impact

(Wilhelms and Squyres, 1984) or several large impacts (Frey and Schultz, 1988) have been suggested as an explanation of the crustal dichotomy. There is, however, no confirming evidence from topography (Smith *et al.*, 1999; Zuber *et al.*, 2000), and no simulations of impacts of such scale have been carried out to test whether the observed pattern of crustal thickness variation (Zuber *et al.*, 2000; Neumann *et al.*, 2004) can be produced. For an endogenic origin of the dichotomy, three different mechanisms have been proposed that are associated: (1) with the evolution of an early magma ocean (Hess and Parmentier, 2001; Elkins-Tanton *et al.*, 2003) (2) with an episode of degree-one mantle convection (Schubert and Lingenfelter, 1973; Wise *et al.*, 1979; Zhong and Zuber 2001), and (3) with an early phase of plate tectonics (Sleep, 1994).

The crustal dichotomy may have arisen from a heterogeneous evolution of a magma ocean or the crystallization of a magma ocean may have led to gravitationally unstable layering (a similar suggestion has been proposed for the Moon; see Section 10.09.6.3), as the late-stage silicates that crystallized at shallow mantle depths were denser than earlier cumulates that crystallized near the base of the magma layer. A potential long-wavelength overturn of an unstable mantle may have thickened the crust over the downwelling region and thinned the crust elsewhere (Hess and Parmentier, 2001).

An episode of long-wavelength mantle convection after solidification of an early magma ocean has been suggested for a layered mantle viscosity structure with a viscosity jump of more than a factor 25 in the mid-mantle (Zhong and Zuber, 2001; Roberts and Zhong, 2006). The time required to develop degree-1 convection ranges from 100 Ma to several hundred Ma and is consistent with most assumptions about the timing of the dichotomy formation. It is not clear what would lead to such a viscosity stratification, but a partially molten zone might be possible.

Sleep (1994) proposed that the smooth northern lowlands and the Tharsis volcanoes were produced as an ocean floor and an island arc volcanic chain similar to these features on Earth. More recently, the magnetic lineation patterns on parts of the southern highlands detected by MGS have been interpreted to be the result of plate divergence (Connerney *et al.*, 1999; Connerney *et al.*, 2005). The dating of the magnetic anomalies led the authors to suggest that the potential early plate tectonic regime lasted about 500 Ma. However, there is no striking geological evidence for such a regime on Mars.

The proposed scenarios for the formation of the crustal dichotomy have different implications for the thermal evolution of Mars, in particular, for the early evolution. At the present stage and most likely during the last 4 Ga, Mars is and was in the stagnant lid regime with a stable plate on top of the convecting mantle. The first few hundred million years, however, is an uncertain epoch and it is unknown whether the planet was in a phase of plate tectonics that later changed to one plate tectonics or whether the planet never changed its heat transport mechanism and always was in the stagnant lid regime.

The existence of a brief episode of plate tectonics has important consequences for the thermal evolution of the planet. Assuming stagnant lid convection throughout the entire evolution of Mars, most models show a continuously cooling interior. Early models based on constant-viscosity laws predicted that a rheological lithosphere has grown in thickness over time and reached a present-day thickness of about 100–200 km depending on the content of radioactive heat sources (Schubert *et al.*, 1992; Breuer *et al.*, 1993; Spohn *et al.*, 1998). The lower value corresponds to models assuming a chondritic heat source concentration in the Martian mantle and the upper bound corresponds to models assuming a lower heat-source concentration equivalent to those derived from the SNC meteorites. The mantle temperatures decrease rapidly and present-day values are about 1400–1600 K. More recent models using the stagnant lid parametrization show that the Martian mantle may have a substantially thicker rheological lithosphere of about 350–450 km (Grasset and Parmentier, 1998; Reese *et al.*, 1998; Breuer and Spohn, 2006). Furthermore, the present mantle temperature of about 1900–2100 K is much higher than expected from the older models with similar parameter values (**Figure 11**).

In contrast to the stagnant lid convection throughout the entire evolution, an early plate tectonics regime suggests a rapidly cooling interior during the first few hundred million years. As soon as plate tectonics stops, the heat transport becomes inefficient and the interior heats up until the temperature again decreases slowly to its present value that does not differ significantly to the temperature of the models with stagnant lid convection throughout the entire evolution assuming the same reference viscosity (**Figure 11**). The temperature increase also suggests a decrease in the thickness of the elastic lithosphere after cessation of plate tectonics and during a long period of Martian evolution. Such a trend, however,

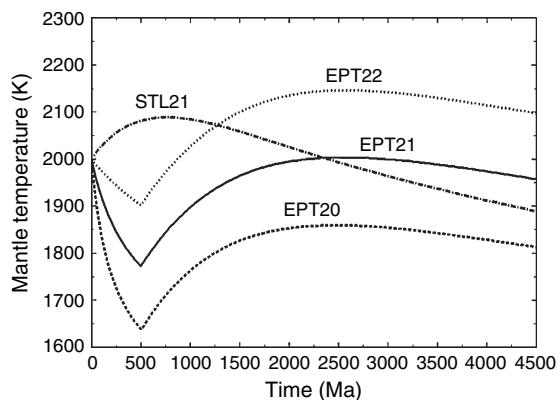


Figure 11 Mantle temperature as a function of time for three models with early plate tectonics. After 500 Ma plate tectonic ceases and stagnant lid convection sets in. Model EPT20 has a reference viscosity of 10^{20} Pa s (dashed line), model EPT21 a reference viscosity of 10^{21} Pa s (solid line), and model EPT22 a reference viscosity of 10^{22} Pa s (dotted line). Also shown is the temperature for the stagnant lid convection model STL21 with a reference viscosity of 10^{21} Pa s (dash-dotted line). This model assumes stagnant lid convection throughout the entire evolution of Mars. Modified from Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics: Evidence from the magnetic field history and crust evolution. *Journal of Geophysical Research – Planets* 108(E7): 5072 (doi:10.1029/2000JE001999).

has not been observed (Zuber *et al.*, 2000; McGovern *et al.*, 2004).

The elastic lithosphere thicknesses which were derived from an analysis of the gravity and topography data suggest a rapid growth during the Noachian and a more modest subsequent growth (Zuber *et al.*, 2000; McGovern *et al.*, 2004). This trend is in general consistent with the thermal evolution models (Schubert *et al.*, 1992; Breuer *et al.*, 1993; Spohn *et al.*, 1998; Breuer and Spohn, 2003, 2006), which show a rapid growth of the elastic lithosphere (defined by an isotherm of 1050 K) from a few kilometers to more than 100 km during the first few hundred million years and later a slower growth to values of 130–200 km. However, the estimated small values of the effective elastic lithosphere thickness of less than 16 km (Zuber *et al.*, 2000; McGovern *et al.*, 2004; Grott *et al.*, 2005, 2007) are nevertheless difficult to explain unless the observed features are Early Noachian in age (more than 4 Ga, Hartmann and Neukum, 2001). In particular the analysis of rift valley topography in the Thaumasia region (Grott *et al.*, 2005, 2007), where the formation of the examined rift zones has actually been dated at the transition of the Noachian to the

Hesperian (approximately 1 Gyr after accretion), suggests an effective elastic lithosphere thickness of about 12 km which is inconsistent with the present models showing a rapid growth of the elastic lithosphere to more than 100 km during the first few hundred million years.

An alternative explanation for the observed rapid growth of the elastic lithosphere thickness is given by the influence of an enriched and low conductivity primordial crust on the thermal evolution (Schumacher and Breuer, 2006). In particular during the early evolution, the mantle underneath this primordial crust is too hot, that is, the 1050 K isotherm is located in the crust (Figure 12), and, therefore, not strong enough to support stresses over a geological timescale. Thus, the rheological behavior of the

crustal composition determines the elastic strength of the lithosphere. As minerals in the crust, for example, quartz, are in general softer than mantle minerals like olivine (Watts, 2001), the isotherm defining the base of the elastic lithosphere is lower and the elastic lithosphere much thinner for the same temperature profile. Assuming, for instance, a diabase crust, the elastic lithosphere thickness is determined by the depth to the 700 K isotherm (Caristan, 1982) and is only a few tens of kilometres which is consistent with the observed effective elastic thicknesses (consider that the effective elastic thickness can be a factor of 2 smaller than the elastic lithosphere thickness). The calculated elastic lithosphere thicknesses can actually be thinner as diabase is a very stiff material for the crust and the isotherm comparatively

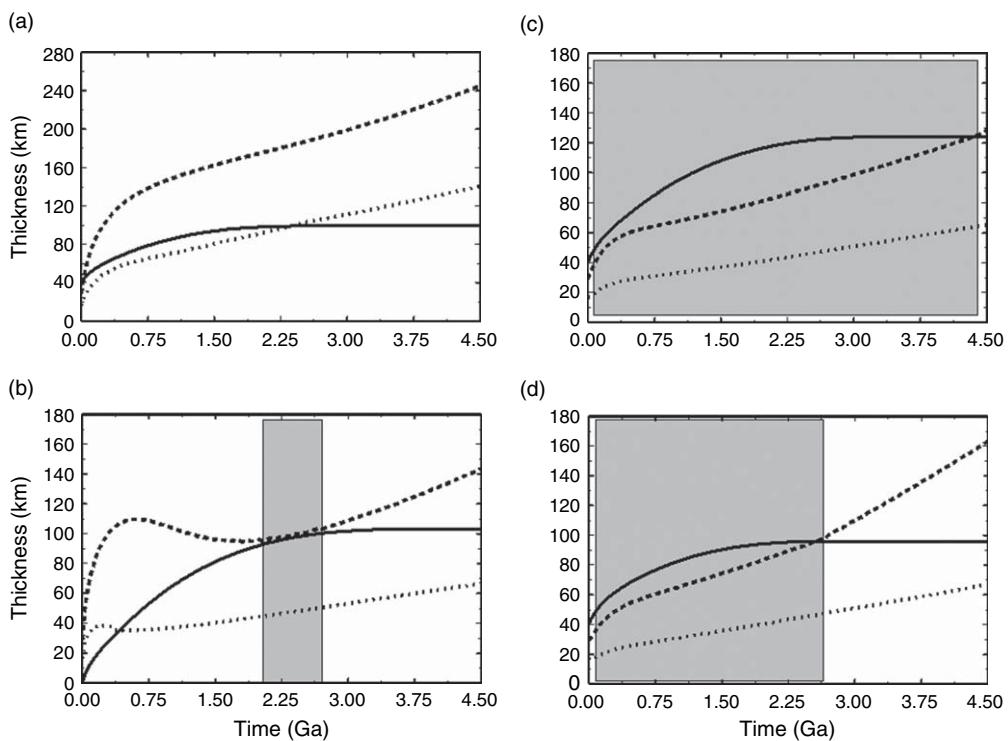


Figure 12 Maximum elastic lithosphere thickness determined by the depth to the 1050 K isotherm representative for olivine-rich mantle material (dashed lines), crustal thickness (solid lines), and depth of the 700 K isotherm representative for diabase crustal material (dotted lines) as a function of time. The results of four models assuming a dry Martian mantle with a reference viscosity of 10^{21} Pas and $k_m(T, P)$ are shown (a) $k_c = 4 \text{ W m}^{-1} \text{ K}^{-1}$, $D_{\text{prim}} = 40 \text{ km}$, and $Q_{\text{dep}} = 20\%$; (b) $k_c = 2 \text{ W m}^{-1} \text{ K}^{-1}$, $D_{\text{prim}} = 0 \text{ km}$, and $Q_{\text{dep}} = 0$; (c) $k_c = 2 \text{ W m}^{-1} \text{ K}^{-1}$, $D_{\text{prim}} = 40 \text{ km}$, and $Q_{\text{dep}} = 20\%$; (d) $k_c = 2 \text{ W m}^{-1} \text{ K}^{-1}$, $D_{\text{prim}} = 40 \text{ km}$, and $Q_{\text{dep}} = 40\%$ with D_{prim} the thickness of a primordial crust and Q_{dep} the initial depletion of the mantle in radioactive heat sources due to primordial crust formation. The shaded areas represent the time when the base of the crust is equal or thicker than the depth of the 1050 K isotherm. During that time only part of the crust but not the mantle provides the necessary strength to support loading. For a crust with a diabase composition the elastic lithosphere is thus given by the depth of the 700 K isotherm. A rapid change of the elastic lithosphere thickness is expected at the interface of the shaded area. Modified from Schumacher S and Breuer D (2006) Influence of a variable thermal conductivity on the thermochemical evolution of Mars. *Journal of Geophysical Research* 111(E2): E02006.

high. As soon as parts of the mantle cool below the olivine isotherm, that is, 1050 K assuming olivine as the dominant mantle mineral, the mantle contributes to the elastic strength and a rapid increase of the elastic lithosphere thickness can be expected if crust and mantle layer are mechanically coupled (Lavier and Stecker, 1997). The time when and whether this transition occurs depends on the efficiency of mantle cooling. A wet mantle and a strong initial depletion of radioactive elements in the mantle cool it more efficiently; therefore, the transition to an elastic lithosphere supported by the strength of the mantle occurs earlier in time. However, the main conclusion here is that such a transition could have occurred sometime during the evolution of Mars but probably during the Hesperian if a sufficiently thick low-conductivity crust existed early in the evolution.

Mars has been volcanically active throughout its history. In fact, the recent detection of young lava flows (<100 Ma) suggests that the planet has been volcanically active up to the recent past (Hartmann *et al.*, 1999). A continuous decline of the volcanic activity has been suggested since the Noachian. Martian volcanism and stratigraphy have been reviewed by Greeley and Spudis (1981) and by Tanaka *et al.* (1992) and Mouginis-Mark *et al.* (1992). The time of crust formation has been constrained using isotope data gathered from SNC meteorites (see Halliday *et al.* (2001) for a recent review) and surface morphological data on volcanic extrusions. A recent study of the Nd mass balance in Martian meteorites (Norman, 2002) proposes a two-stage formation of the Martian crust with 20–30 km crust formed early, possibly simultaneously with core formation, and 45–75 km thereafter from depleted mantle sources. The exact timing of the second stage, however, cannot be constrained with the geochemical data.

The present thickness of the Martian crust is also not very well known and estimates of the average thickness vary between 40 and 150 km. The upper value has been derived from models of the interior structure of Mars based on the MoI factor and the chemistry of the SNC meteorites (Sohl and Spohn, 1997; Sohl *et al.*, 2005) and the lower value from an Bouguer inversion of the gravity and topography data assuming Airy isostasy (or a crust density that is everywhere the same) and further assuming that the floor of the Hellas impact basin defines the minimum crust thickness (Zuber *et al.*, 2000) (see Chapters 10.05 and 10.02 for more details).

The thermochemical evolution models that include crustal growth usually predict that most of the crust formed in the first billion years and had grown to reach a thickness of some tens of kilometers (Spohn, 1991; Schubert *et al.*, 1992; Breuer *et al.*, 1993; Hauck and Phillips, 2002; Breuer and Spohn, 2003, 2006) or a few hundreds of kilometers (Weizman *et al.*, 2001). These models in general confirm the observation of a continuous decline of global volcanism and crustal growth since the Noachian. In contrast, models with early plate tectonics in Mars (Breuer and Spohn, 2003) show an early efficient cooling during the phase of plate tectonics that will either frustrate later crust production and may not explain, for example, the northern volcanic plains in the Early Hesperian or show a crustal growth rate with a late peak (about 2 Ga) which is not observed.

The strict assumption that most of the crust is produced during the first few hundred million years (Solomon *et al.*, 2005; Nimmo and Tanaka, 2005), places a strong constraint on thermal evolution models that assume stagnant lid convection throughout the entire evolution. Hauck and Phillips (2002) argue that only a wet mantle with an initial mantle temperature of about 1700 K can explain this observation. In contrast to these findings, Breuer and Spohn (2006) show that models with a dry Martian mantle are also consistent with the crustal evolution if the initial mantle temperature after core formation is high enough and if a primordial crust has been formed during core formation (Figure 13).

It has not been established whether these models can also explain the observed recent volcanic activity in the large volcanic provinces, Tharsis and Elysium (Hartmann *et al.*, 1999; Berman and Hartmann, 2002; Neukum *et al.*, 2004). In these regions the volcanic activity was most likely active for a few billion years although the bulk of the volcanic provinces was present since the Noachian (Banerdt and Golombek, 2000; Phillips *et al.*, 2001). The thermochemical evolution models (Hauck and Phillips, 2002; Breuer and Spohn, 2003, 2006) suggest that crustal formation ceased 1–3 billion years ago based on the fact that the average mantle temperature fell below the mantle solidus. The existence of longstanding volcanism in Tharsis and Elysium up to recent times could be explained with stable and longstanding plumes that have a substantially higher temperature (i.e., more than 100 K) than the average mantle. Possible mechanisms to generate stable and longstanding plumes may be due to a deep phase transition (Weinstein, 1995; Harder and Christensen, 1996;

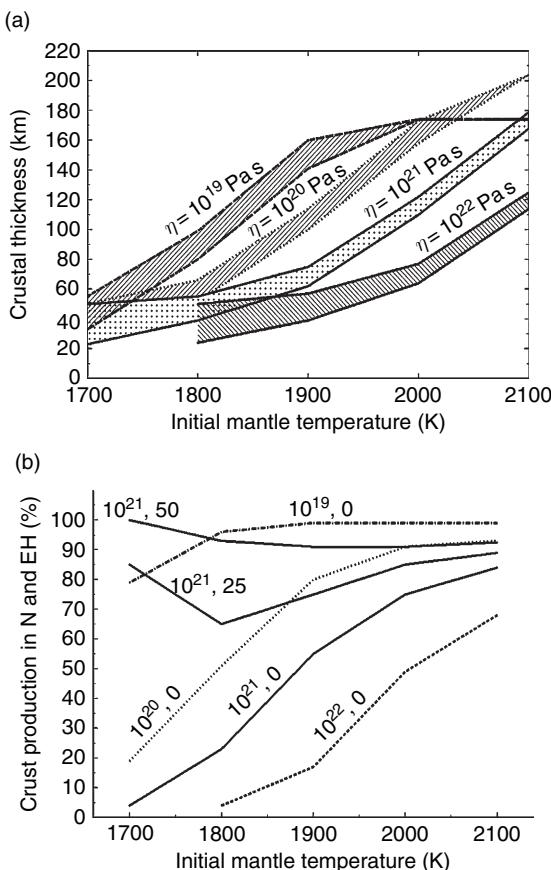


Figure 13 Present-day crustal thicknesses as a function of the initial mantle temperature (a). Models have been calculated for reference viscosities η_{ref} between 10^{19} and 10^{22} pas. The upper bounds of the individually hatched areas represent models with a primordial crust of 50 km thickness and the lower bounds represent models without any primordial crust. Percentage of total crust production during the Noachian and Early Hesperian (b). The models are labeled by the assumed reference viscosity in Pas and the thickness of the primordial crust in km. Modified from Breuer D and Spohn T (2006) Viscosity of the Martian mantle and its initial temperature: Constraints from crust formation history and the evolution of the magnetic field. *Planetary and Space Science* 54: 153–169.

Breuer *et al.*, 1998; Harder, 2000). In particular, the endothermic spinel to perovskite phase transition that might be located close to the CMB (see also Chapter 10.02) helps to stabilize the plumes and reduces their number to one or two upwellings. Calculations with phase changes fail to produce a one-plume structure with a more realistic temperature-dependent viscosity unless the activation energy is unreasonably low ($<60 \text{ kJ mol}^{-1}$) (Roberts and Zhong, 2004). Another mechanism to generate longstanding plumes has been proposed assuming chemical layering either

with a variable mantle thermal conductivity (Schott *et al.*, 2001) or with the presence of the crustal dichotomy (Wenzel *et al.*, 2004). For these mechanisms, the existence of a chemically layered mantle is required either due to fractionation of a magma ocean (Wenzel *et al.*, 2004) or as the consequence of crustal formation due to the formation of a depleted upper mantle layer, which does not remix with the remaining undepleted mantle (Schott *et al.*, 2001). An alternative to generate melt without the existence of a large plume underneath the volcanic regions has been proposed by Schumacher and Breuer (2006). Once the main structure of the volcanic provinces has been built up, the lateral thickness variations of the low-conductivity crust allow the continuous generation of melt underneath these thickened regions and may explain the longstanding and recent volcanism in Tharsis and Elysium (Figure 14). It has been suggested that the crust underneath Tharsis is about 20–40 km thicker than in its surroundings (Neumann *et al.*, 2004).

At present, Mars has no global dipole magnetic field, but it is conceivable that it generated a field in the iron-rich core early in its history (see Chapter 10.07). This assumption arises from the detection of the strongly magnetized ancient crust on Mars; one of the most astonishing outcomes of the MGS mission (Acuña *et al.*, 1998; Acuña *et al.*, 1999; Connerney *et al.*, 1999; Acuña *et al.*, 2001). After the MGS orbit injection in 1997 and during the aerobraking phase and science-phasing orbit phase, it was possible to get magnetic measurements at periapses ranging from 85 to 170 km above Mars' surface. Connerney *et al.* (1999) (see also Purucker *et al.*, 2000; Acuña *et al.*, 2001) have discovered strongly magnetized regions in the crust, closely associated with the ancient, cratered terrain of the highlands in the southern hemisphere. There has been a great interest in the meaning of the spatial pattern of magnetization (Jurdy and Stefanick, 2004; Arkani-Hamed, 2004), including possible lineations (these crustal magnetizations even suggest reversals) that were believed to suggest an analogy to plate tectonics (Connerney *et al.*, 1999, 2005), but other models favor thermal remanence acquired during the time these ancient rocks cooled in conjunction with dike intrusions (Nimmo, 2000). The inferred crustal magnetizations, which are up to $\sim 10\text{--}30 \text{ A m}^{-1}$, are an order of magnitude stronger than those of the Earth's continents (Toft and Arkani-Hamed, 1992; Arkani-Hamed and Dyment, 1996) and comparable in magnitude with the remanent magnetization of fresh extrusive basalt near the oceanic ridge axes of

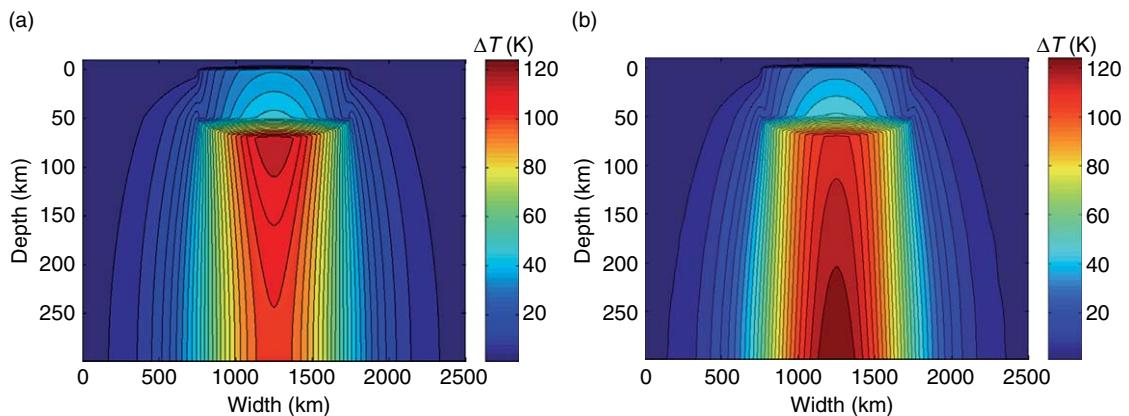


Figure 14 Temperature differences between a model with and a model without crustal thickening (absolute temperature increase caused by a 20 km thickening of a low-conducting crust) for (a) constant mantle thermal conductivity $k_m = 4 \text{ W m}^{-1} \text{ K}^{-1}$ and (b) k_m temperature- and pressure-dependent, respectively. The main difference between the two models is that the temperature and pressure dependence of the thermal conductivity in the mantle results in a maximal temperature increase at the bottom of the stagnant lid instead below the crust for constant k_m . This circumstance helps to generate partial melt preferentially at the bottom of the stagnant lid below thickened crust where fresh mantle material can be refilled by mantle convection.

Earth (Bleil and Petersen, 1983). These magnetic anomalies indicate the existence of a strong ancient intrinsic Martian magnetic moment corresponding to a magnetic field strength of 0.1–10 times that of the Earth’s at present (Ness *et al.*, 1999; Mitchell *et al.*, 1999, 2001).

There are several ways in which the Martian crust could have been magnetized (*see* Chapter 10.07). One of the most probable ways is thermoremanent magnetization (TRM). With TRM, the magnetization is produced when rock cools below a critical temperature (Curie temperature) in the presence of a magnetizing field. This is an effective mode for producing an intense remanent field. The magnitude of magnetization produced in the rock depends on the strength of the internal field, the mineralogy, and the magnetic microstructure. There is a tradeoff between the concentration of magnetic carriers and the strength of the magnetic field: The lower the magnetic field the more magnetic carriers are required to explain an observed magnetization. Assuming an early Martian magnetic field similar to the present-day Earth field, the concentration of magnetic carriers might be comparable to the amount in extrusive basalt. Bearing in mind the possible reduction of the remanent magnetization through viscous decay and chemical alterations during this very long geologic time (the magnetization of the oceanic basalt decreases by a factor of 4–5 during the first 20 My, largely by chemical alterations (Bleil and Petersen, 1983)), the initial magnetization of the Martian source bodies must have been even stronger.

However, there is ample evidence that the FeO content of the Martian mantle is about twice that of the Earth’s mantle (Sohl and Spohn, 1997; Sanloup *et al.*, 1999). Whether this high concentration of FeO translates to a high concentration of magnetic minerals depends on the oxidation state of the Martian mantle and lower parts of the crust. However, for a high concentration of magnetic minerals, the early magnetic field could have been smaller than that of the present Earth.

The timing of the dynamo places strong constraints on the thermal evolution of the planet. It has been proposed on the basis of the surface distribution of the magnetic anomalies that this magnetization event predated the formation of the Hellas and Argyre basins (Connerney *et al.*, 1999), roughly at the end of the Early Noachian (Head *et al.*, 2001). The existence of an early magnetic field has been supported by the ancient Martian meteorite ALH84001, which shows a magnetization that was acquired at 4 Ga or even earlier (Weiss *et al.*, 2002). It has been alternatively suggested that the dynamo turned on after the giant impact craters formed implying that the remanent magnetization of the crust was acquired later in Martian history (Schubert *et al.*, 2000). This hypothesis is supported by the SNC meteorites with formation ages of 1.3 Ga–180 Ma, which show a remanent magnetization consistent with a weak ancient surface field of a few hundred to a few thousand nanoteslas (Collinson, 1997; Rochette *et al.*, 2001). This weak surface field may suggest a late onset or reactivation of a core

dynamo that would have stopped operating before the present. However, it is also possible that the young Martian meteorites were magnetized by the older and strongly magnetized crust rock (see Connerney *et al.* (2003) for a review).

The evidence of early dynamo action is supported by most thermal evolution models that consider the magnetic field history (e.g., Schubert and Spohn, 1990; Spohn *et al.*, 1998; Hauck and Phillips, 2002; Breuer and Spohn, 2003, 2006; Williams and Nimmo, 2004). The models assume a thermally driven dynamo during the first few hundred million years as the consequence of the early rapid cooling if the core is superheated with respect to the mantle (Figure 15). In the subsequent evolution the core remains liquid with a heat flow below the critical value and, thus, no dynamo action until today. A totally molten core for the entire evolution is possible if the core contains at least ~ 5 wt.% sulfur (Williams and Nimmo, 2004). A core sulfur abundance of 14.2 wt.% is suggested for Mars from geochemical analyzes of the SNC meteorites (Dreibus and Wänke, 1985, McSween, 1985). Assuming a sulfur content in the core below 5 wt.%, inner-core growth and consequently a chemical dynamo would be possible. Such a scenario, however, is at variance with the assumption of a brief, very early dynamo. If core solidification did take place on Mars, it likely would have generated a very long-lived dynamo (Schubert *et al.*, 1992).

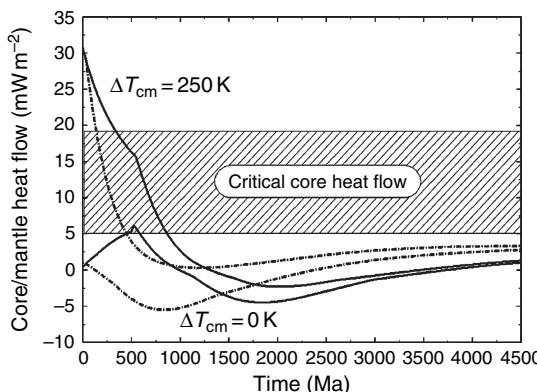


Figure 15 Core–mantle heat flow as a function of time for models with early plate tectonics (solid line) and models with stagnant lid convection throughout the entire evolution (dash-dotted line) with initial temperature differences across the CMB of $\Delta T_{cm} = 0$ K and $\Delta T_{cm} = 250$ K. Modified from Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics: Evidence from the magnetic field history and crust evolution. *Journal of Geophysical Research – Planets* 108(E7): 5072 (doi:10.1029/2000JE001999).

An alternative scenario assumes a change in the heat transport mechanism of Mars as suggested by the model of early plate tectonics (Nimmo and Stevenson, 2000; Stevenson, 2001). During the phase of plate tectonics the interior cools efficiently and a thermal dynamo but also a chemical dynamo is possible if the core cooled rapidly below its melting temperature. As soon as this regime is followed by stagnant lid convection, the mantle heats up and dynamo action stops until today. In the case of a thermal dynamo, the heat flow decreases below the critical value (Figure 15). In the case of a chemical dynamo, the end of the plate tectonic regime marks the time of the coldest interior (Figure 11). Thus, in the subsequent evolution the core temperatures remain above the melting temperature and inner-core growth and chemical dynamo action stop.

10.09.7 Io

10.09.7.1 Tidal Heating and the Evolution

Io is unique in the solar system in that its primary source of internal heat is not radioactive decay but tidal dissipation (Peale *et al.*, 1979). Io is the Galilean satellite nearest to Jupiter (see Chapter 10.13 for a discussion on the giant planets) and it is the most volcanically active body in the solar system. Unlike most bodies in the solar system, which would require *in situ* methods to determine the heat flow, Io's surface heat flow can be estimated from remote observations of its thermal emission. From a global inventory of Io's thermal radiation from a variety of sources, McEwen *et al.* (2004) obtained a global heat flow estimate of 2.1 ± 0.7 W m⁻², while Veeder *et al.* (2004) give 3 ± 1 W m⁻². The true value most likely falls in the range from 1.5 to 4 W m⁻² (60–160 TW total). The heat production rate implied by the surface heat flow far exceeds that which can plausibly be produced by radioactive decay; thus, a different internal heat source must be active. Io's location so close to Jupiter causes it to be subject to intense tidal forces. The constantly varying tidal potential resulting from Io's eccentric orbit results in more than 10 m radial deformation of Io's solid surface every orbit (1.8 days). Viscoelastic dissipation of heat in the interior of Io due to the repeated tidal deformation is responsible for Io's large internal heat source. The tremendous heat flow measured at the surface of Io is the result of silicate volcanism, with typical inferred eruption temperatures of 1200–1400 K, and extreme temperatures over 1800 K (McEwen *et al.*, 1998). It is

likely that the sulfur-based volcanism observed by Voyager is also driven by silicate magma.

Despite the intensity of the volcanic activity on Io and the inferred high temperatures of its interior, Io supports some of the highest and steepest topography of any planetary body. Mountains up to 17 km high dot the surface (Schenk *et al.*, 2001; Jaeger *et al.*, 2003), and, oddly, these are not volcanoes. Instead, volcanic centers are generally depressions, while the mountains appear to be fault-bounded tectonic uplifts. The presence of such large topography seems incompatible with a lithosphere that must allow the transport of $\sim 2.5 \text{ W m}^{-2}$ of heat.

A solution to this problem was found by O'Reilly and Davies (1981), who proposed a heat-pipe mechanism for heat transport through Io's lithosphere accomplished by melt transport through fissures. The most important implication of this model is that, as successive eruptions spread across the surface and cool, they bury the flows of previous eruptions, resulting in an advective transport of cold material from the surface downward. In order to match the observed heat flow, Io must be resurfaced globally by silicate magma to a depth of 1–2 cm every year. This is then the rate at which material is advected downward within the lithosphere. The equation for the temperature T in the lithosphere is then

$$\frac{k}{\rho C_p} \frac{d^2 T}{dz^2} = v \frac{dT}{dz} - \frac{H}{\rho C_p} \quad [23]$$

where k is the thermal conductivity, ρ is the density, C_p is the specific heat, z is the depth from the surface, v is the downward velocity (resurfacing rate), and H is the volumetric heat production. The temperature must match the surface temperature T_s at the surface and the melting temperature T_m at the base of the lithosphere. The solution in the case of no heat production is

$$T(z) = T_s + \frac{(T_m - T_s)(e^{\lambda \xi} - 1)}{e^\lambda - 1} \quad [24]$$

where $\xi = z/D$ is the depth normalized by the thickness of the lithosphere D , and the dimensionless parameter $\lambda = DvC_p/k$ is the advective velocity normalized by the conductive velocity scale. This can be related to the heat flux carried by melting F as follows:

$$\lambda = \frac{DC_p F}{k[L_f + C_p(T_m - T_s)]} \quad [25]$$

where L_f is the latent heat of fusion of the silicate rocks. For a heat flux of 2.5 W m^{-2} and a 30-km-thick

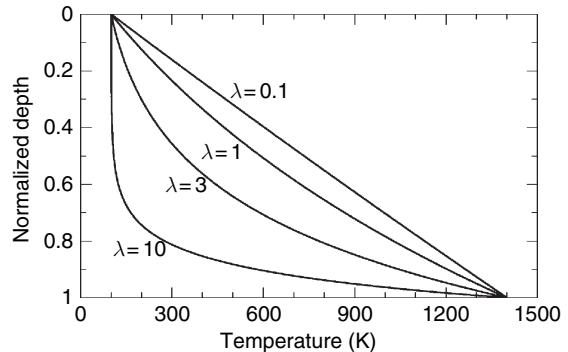


Figure 16 Temperature as a function of normalized depth for different values of λ , the normalized advective velocity defined in the text. Modified from Moore W, Schubert G, Anderson JD, and Spencer JR (2006) The interior of Io. In: Lopes RMC and Spencer JR (eds.) *Io after Galileo: A New View of Jupiter's Volcanic Moon*, pp. 89–108. Chichester, UK: Springer and Praxis Publication. With kind permission from Springer Science and Business Media.

lithosphere, λ is about 10. Solutions for different values of λ are shown in Figure 16. For λ of 3, more than 80% of the thickness of the lithosphere is below 900 K, and for λ of 10, more than 95% of the lithosphere is cold enough to sustain elastic stresses for very long periods of time. This is how Io's lithosphere can support huge mountains while at the same time allowing a heat flux of 2.5 W m^{-2} to pass through.

What sort of melt fractions does the heat pipe mechanism imply for the asthenosphere? It is straightforward to calculate the melt segregation velocity due to Darcy flow driven by the buoyancy of the melt (e.g., Scott and Stevenson, 1986):

$$\Phi v = \frac{k_\Phi g \Delta \rho}{\eta_m} \quad [26]$$

where Φ is the melt volume fraction (porosity), g is the gravitational acceleration, $\Delta \rho$ is the difference between the solid and melt densities ($\sim 500 \text{ kg m}^{-3}$), η_m is the melt viscosity ($\sim 1000 \text{ Pa s}$), and k_Φ is the permeability, which is related to the porosity by a function of the form

$$k_\Phi = \frac{b^2 \Phi^n}{\tau} \quad [27]$$

where b is a typical grain size ($\sim 1 \text{ cm}$), and n and τ are constants which are functions of the geometry of the melt. The dependence of k_Φ on grain size is overly simplified in this model, since real systems may have broad grain size distributions. Using the 1 cm yr^{-1} resurfacing velocity (which is actually the melt flux Φv) and inserting experimentally

determined values for the constants n [3] and τ (200) (Wark and Watson, 1998; Liang *et al.*, 2001), results in an estimate for Φ between 10% and 20% (Moore, 2001). Melt fractions exceeding this will transport too much heat (the heat transported goes as k_Φ), thus cooling the asthenosphere and bringing the melt fraction back to the equilibrium value.

A rough estimate for the thickness of the asthenosphere may be obtained by extrapolating along the adiabat from the temperature at the top of the melt zone to the solidus:

$$d = \frac{\Delta T_m}{(dT/dz)_m - (dT/dz)_{ad}} \quad [28]$$

where $\Delta T_m = \Phi L_f / C_p$ is the temperature excess required to reach a melt fraction Φ and the denominator is the difference between the slopes of the solidus ($\sim 0.8 \text{ K km}^{-1}$) and the adiabat ($\sim 0.1 \text{ K km}^{-1}$). Using the melt fractions estimated above, the asthenosphere is 60–120 km thick. Below this is a mantle which is solid, but quite close to the solidus. Though we lack direct evidence, the core is most likely liquid, since the iron solidus is below the silicate solidus. This simple analysis of melt segregation ignores the role of convection in heat transport within Io. Despite its broad applicability to planetary heat transport, however, heat transport in partially molten, convecting systems has not received thorough quantitative treatment. Unlike subsolidus convection, for which detailed parametrizations exist which incorporate many aspects of the physics of planetary materials (particularly, the rheology of rocks), partially molten convection has been treated in an *ad hoc* fashion by tacking a melt segregation analysis such as that presented above onto a subsolidus convection parametrization. This approach ignores the strong influence of melting and melt segregation on the convective flow itself, which is particularly important when the rheology of the convecting layer influences the heat production as it does in a tidally heated body such as Io.

10.09.7.2 Thermal and Orbital Evolution

The tidal heat source is not only capable of much greater heating rates than the radioactive heat sources, but it also evolves with the orbit of Io, since it depends on both the distance of Io from Jupiter (the semimajor axis) and the eccentricity of Io's orbit. Tidal heating is also very sensitive to the rheology of Io's interior. The heat source is therefore coupled to both the orbit and

the interior temperature. This results in a very different thermal evolution for Io than for other bodies of similar size (e.g., the Moon).

The tidal heating of Io's interior is a result of the resonant orbital interactions between Io, Europa, and Ganymede, which have orbital periods that are in the ratio $\tau_E : \tau_E : \tau_G = 1:2:4$. The orbits of all three satellites have evolved together into what is called the Laplace resonance, where the conjunctions of each satellite with its nearest outward neighbor occur when the inner satellite is at perijove (nearest to Jupiter) and the outer satellite is at apojove (furthest from Jupiter), thus maximizing their separation. These repetitive alignments cause the orbits to remain elliptical, even though dissipation of tidal energy in the satellites should tend to circularize their orbits (while dissipation in Jupiter tends to make them more eccentric). In addition to the effects of dissipation on eccentricity, tides raised on the satellite tend to reduce the semimajor axis of the orbit, that is, the satellite tends to spiral inward. Thus, Io's orbital evolution is tightly coupled to the thermal evolution of its interior. Conversely, tides raised on Jupiter by the satellite cause the satellite to spiral outward (like the Moon). Thus Io's actual rate of orbital migration represents a balance between dissipation in Io and dissipation in Jupiter. Indeed, this orbital migration is thought to be the process by which the resonant configuration is assembled (Yoder, 1979).

The coupled thermal-orbital evolution of Io (and Europa and Ganymede) is a complex dynamical system with a wide possible range of behaviors. The interior temperature is set by the balance between tidal heating and heat transport (convective or magmatic). Orbital evolution is driven by dissipation in Io. Fischer and Spohn (1990) studied the coupled system including a simplified orbital evolution (linearized about the Laplace resonance) and convective heat transport, and recognized that there are multiple equilibria for the thermal state of Io, as illustrated in **Figure 17**, a cartoon depicting the possible thermal equilibria in a tidally heated body. Radioactivity (solid line) produces the same heating regardless of temperature, while tidal heat production (solid curve) depends strongly on temperature, increasing with temperature until the Maxwell time (viscosity over shear modulus) of the material approaches the tidal forcing period. At a critical temperature (T_c) slightly above the solidus the shear modulus begins to decrease rapidly (Berckhemer *et al.*, 1982), resulting in a drop in tidal heating. At the breakdown temperature T_b , the solid matrix loses coherence, and the material becomes dominated by the mechanical properties of the liquid

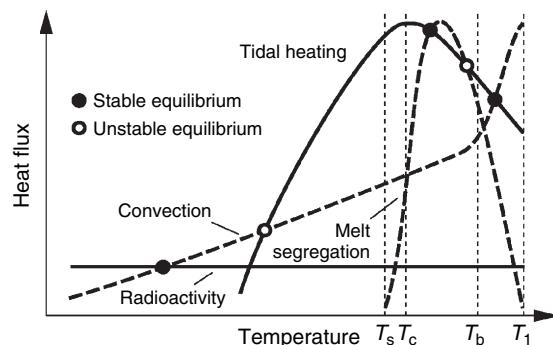


Figure 17 Cartoon depicting the possible thermal equilibria (circles) in a tidally heated body. Heat sources (radioactivity and tidal heating) are shown as solid lines, and heat transport mechanisms (convection and melt segregation) are shown as dashed lines. Also marked schematically are the solidus T_s , critical T_c , breakdown T_b , and liquidus T_l temperatures. Modified from Moore W, Schubert G, Anderson JD, and Spencer JR (2006) The interior of Io. In: Lopes RMC and Spencer JR (eds.) *Io after Galileo: A New View of Jupiter's Volcanic Moon*, pp. 89–108. Chichester, UK: Springer and Praxis Publication. With kind permission from Springer Science and Business Media.

and the tidal heating is greatly reduced. Convective heat transport (dashed curve) increases with temperature, as the viscosity decreases, becoming extremely efficient as the liquidus temperature T_l is approached. Melt segregation (as described above) can only occur between the solidus and liquidus, and is not efficient above the breakdown temperature (since there is no longer any solid matrix for the melt to segregate from). Depending on the details of the heat production and transport processes, not all of these equilibria may be realized, so we use a qualitative diagram to illustrate the possible range of behaviors.

Fischer and Spohn (1990) identified the highest and lowest temperature equilibria (the points where the solid tidal heat and radioactivity curves intersect the dashed convection curve) and noted that at times of low eccentricity, the high-temperature equilibrium may cease to exist. This is because low eccentricity shifts the tidal heat production curve downward, and it may fail to intersect the convective heat transport curve at all. Such episodes are followed by rapid cooling, as Io evolves toward the more common radioactivity-driven convective equilibrium. Moore (2003) investigated the tidal-convective equilibrium (for the current eccentricity, independent of the orbital evolution) and concluded that the high-temperature equilibrium, while present, cannot explain the observed heat flow, thus a different heat transport process (melt segregation) was necessary.

Melt segregation can introduce additional equilibria as shown in Figure 17. These equilibria are necessarily between the solidus and liquidus temperatures (T_s and T_l). The simplified calculation given in Section 10.09.7.1 required melt fractions of 10–20% to achieve equilibrium. This equilibrium is stable, but may disappear at times of low eccentricity. Since this equilibrium is reached at temperatures lower than the convective equilibrium (this is a consequence of the high viscosity of rocks, even when partially molten), it is likely that Io never reaches the high-temperature convective equilibrium. Io's volcanic activity and high heat flow strongly suggest that Io is in (or near) an equilibrium state with melt segregation balancing tidal heat production. It is not known whether our roughly three decades of heat flow observations are representative of the long-term average, but if 10^{14} W is the rate of energy dissipation in Io, then if we know the rate of dissipation in Jupiter, we can predict the rate at which Io's orbit is evolving. Unfortunately, the mechanisms by which tidal motions are dissipated in a deep atmosphere such as Jupiter's are not understood at present. We are therefore left with trying to measure the orbital evolution to estimate dissipation in Jupiter. Recent attempts to measure the rate of Io's orbit migration using astrometric measurements of satellite mutual events (eclipses) disagree on both the rate and the direction of the orbital migration. The process used to infer the migration rate of Io from the mutual event timing is apparently quite model dependent, since Lieske (1987) derives a rate of change of the mean motion of Io \dot{n}/n of $-0.74 \pm 0.87 \times 10^{-11} \text{ yr}^{-1}$, while a more recent measurement using a different orbital theory by Aksnes and Franklin (2001) yields a rate of $+36 \pm 10 \times 10^{-11} \text{ yr}^{-1}$. We can put certain bounds on the total outward migration of Io's orbit (assuming Europa and Ganymede have moved in concert) by assuming Io formed as close to Jupiter as possible ($2.4 R_J$) and evolved to its current orbital distance over the age of the solar system. This upper bound on the migration rate, however, is still much too low to account for Io's current dissipation rate, which depends on the maintenance of a high eccentricity (and thus high dissipation in Jupiter) (Yoder and Peale, 1981). We are therefore forced to conclude that Io's orbital evolution has not been steady over the age of the solar system, or that the current heat flow does not represent a long-term average of the tidal dissipation. Either case is possible, since we do not know if the Laplace resonance is primordial or if it was assembled more recently by orbital

migration. Furthermore, Io's surface heat flow represents dissipation integrated over some unknown transport time. However, if most of the heat is carried by melt segregation, as seems to be the case (above and Moore (2003)), the transport time is short (tens of years), and the heat flow does reflect recent tidal dissipation. A longer transport time regulated by convective motions would allow for the possibility of oscillatory states in which the heat flow and dissipation are out of phase (Fischer and Spohn, 1990; Hussmann and Spohn, 2004).

10.09.8 Summary

This chapter has reviewed the study of the interior structure, chemistry, and dynamics of terrestrial planetary bodies: Mercury, Moon, Venus, Mars, and Io. There are many processes which connect these aspects of planetary interiors, many of which have been discussed here. Fundamentally, the dynamics of terrestrial bodies are driven by the transport of heat from the interior to the surface. Heat generation and transport are the processes which connect the structure and chemistry of a planetary body with its dynamical behavior, for example, volcanism, tectonism, and magnetism. The physical and chemical properties of planetary materials influence the mechanics of the planetary heat engine driving dynamics, while the dynamics in turn alter the physical and chemical properties of the planet's interior.

The density structure of each of the terrestrial bodies is roughly similar, with silicate mantles surrounding metallic cores, but the geometries vary significantly, from shallow mantles (Mercury) to very small cores (Moon). The aspect ratio of the mantle can have a strong influence on the planform of convective motions in weakly convecting systems. The silicate mantles are similarly differentiated into crust and mantle portions, with considerable variation in the thickness of the crust. The process of crustal formation, while universal, has not been identical on each of the terrestrial bodies. Both primordial (Moon, Mercury) and evolved (Venus, Io) crusts exist today. This likely reflects the importance of crustal recycling and continuous crustal formation on the more active bodies. Crustal differentiation may have several different effects on the planetary heat engine. First, as a form of heat transport, melt segregation can be extremely efficient as seen at Io. For less-active bodies, the crust can act both as an insulator for the mantle and as a sink for heat-producing

elements. These effects have opposite influences on the temperature of the underlying mantle, the former tends to raise the temperature while the latter depresses it. Crustal material which is unevenly distributed may influence the temperature distribution in the upper mantle and the planform of convection, perhaps explaining the distribution of volcanism on Mars. Finally, melt segregation from the mantle may also be associated with devolatilization of the interior, which influences the rheology of the remaining mantle material.

The thermal evolution of the terrestrial planetary bodies through the history of the solar system has been studied by using parametrizations of heat transport through their mantles. Many physical processes have been incorporated in these models, including complex rheology (with temperature, pressure, stress, and composition dependence), partial melting, crust formation, core heat flux, and inner-core growth. These models are compared with observations of crustal and lithospheric properties, magnetic fields, tectonism, and volcanic activity. Significant uncertainties in the properties of the interiors of the terrestrial planets remain, so only general conclusions can be reached.

Mercury's magnetic field, if produced by dynamo action in the core, provides a very strong constraint on the heat transport through the mantle. If freezing of an inner core is required for dynamo action, this also places a constraint on the sulfur content of the core. Mercury's surface also records a few kilometers of radial contraction over its history in the form of lobate scarps. In order to generate a present-day magnetic field without completely freezing the core, convection in Mercury's mantle must be very weak or possibly have ceased altogether. A stiff rheology and a strong depletion in potassium seem to be required.

The situation on Venus is less well constrained, in part because much of the evidence of Venus' early history has been erased from the surface. Venus has been recently active (within the last 108 years), but the cratering record is permissive of a wide variety of resurfacing scenarios. The strong correlation of gravity and topography at low harmonic degrees indicates that there are large-scale density anomalies in the mantle, likely the result of mantle convection. Whether Venus ever experienced a plate-tectonic phase (or even several) is uncertain, but there is little present evidence for active plate tectonics. The observation that Venus lacks a magnetic field may

be related to the relatively low heat transport efficiency of a single-plate planet with a rigid lid.

Unlike the more active bodies, which largely erase the influence of their initial conditions, the present state of the Moon may be strongly reflective of its formation. The thick lunar crust appears to be a primordial differentiate from a magma ocean that formed during the accretion of the Moon from a hot debris disk orbiting the Earth after a giant impact. The differentiation of the magma ocean may have led to an initially unstable density stratification that drove an early episode of convection, but the segregation of heat-producing elements out of the mantle may result in only weak convective motions and fairly little internal cooling since then. The lunar core is small (<500 km), and its composition is unknown. Certain lunar rocks have been magnetized, but whether this is evidence of an internally generated magnetic field is not certain. If there was an early lunar dynamo, it seems to have gradually decayed with time, beginning at about 4 Ga and ceasing at about 3 Ga. This perhaps reflects an early period of vigorous convection, but some mechanism is required to prevent dynamo onset earlier than 4 Ga unless the earliest remanent magnetizations are not measurable or existent although an early dynamo action was active.

Our knowledge of the interior and the evolution of Mars has been vastly improved by recent satellite and lander missions. The size of the Martian core is now better constrained, as is the average crustal thickness. Due to the ambiguity in these parameters as we do not know the composition and density of core, mantle, and crust, only future seismic experiments may further improve our knowledge. Like the Moon, Mars has ancient, heavily cratered highlands, but unlike the Moon, the less-cratered lowlands are not clearly impact basins. The ancient highlands, however, are regions of thickened crust, suggesting a primordial origin, though the mechanism and exact timing of the dichotomy formation is still debated. Geochemical evidence from SNC meteorites suggests that little mixing has gone on since early in Mars' history and thus that the bulk of the Martian crust was formed very soon after accretion. Mars has been volcanically active into the recent past, and there are a wide range of tectonic features mostly related to volcanic construction. Discovery of strong magnetic anomalies associated with magnetized crust has led to new models of the early thermal history of Mars that include an episode of plate tectonics or an initially superheated core due to its formation to

provide the required core cooling. Since that epoch, the activity of the Martian interior has been gradually declining. The crustal dichotomy provides a spatially variable boundary condition to the convecting mantle and may provide a means for keeping some partial melt and for focusing it at Tharsis.

Although it is an outer planet satellite, Io is essentially a terrestrial body, with a silicate mantle and a metallic core. Io is unique, however, in that its internal heat source is tidal dissipation rather than radioactive decay. This ties the heat production in the interior to the internal physical state through the rheology. Io is also unique in that the dependence of tidal heating on the material properties is such that the internal temperature is kept above the solidus. This leads to massive amounts of volcanic activity and makes Io the only terrestrial body to transport heat primarily by melt segregation. Io today may provide us an opportunity to study processes that were in effect on the other terrestrial bodies only briefly in their earliest histories. The constant resurfacing by volcanic eruption has erased all evidence of impacts on Io's surface and produces strong compressional stresses as the surface subsides beneath subsequent flows. These stresses lead to large fault-bounded massifs, some of the tallest in the solar system. Ionian volcanoes, however, are broad, low-relief structures, perhaps because the abundant sulfur deposits on the surface and in the crust are easily weakened by the locally high heat flow.

The study of terrestrial planet interiors has advanced significantly in the last decades, driven by improvements in modeling, increasing computing power, and new observations of Venus, Mars, the Moon, and Io. The coming decades will see missions to Mercury, the Moon, Mars, and the outer solar system that will expand our knowledge of terrestrial bodies. Extrasolar terrestrial planets may also be detected, allowing us to put the formation of our own planetary system in a broader context. Our understanding of the properties and behavior of planetary materials, especially at high temperature and pressure, will continue to grow through experimentation and observation. Looking further ahead, there will come a day when terrestrial planetary bodies will be directly explored by human beings, and a new era of terrestrial planet geology will dawn. Though that day may be far in the future that achievement will build upon the research summarized in this chapter, extending and refining our understanding of the terrestrial planets.

References

- Acuña MH, Connerney JEP, Ness NF, et al. (1998) Magnetic field and plasma observations at Mars: Initial results of the Mars Global Surveyor Mission. *Science* 279: 1676–1680.
- Acuña MH, Connerney JEP, Ness NF, et al. (1999) Global Distribution of crustal magnetism discovered by the Mars Global Surveyor MAG/ER Experiment. *Science* 284: 790–793.
- Acuña MH, Connerney JEP, Wasilewski P, et al. (2001) Magnetic field of Mars: Summary of results from the aerobraking and mapping orbits. *Journal of Geophysical Research* 106(E10): 23403–23418.
- Aksnes K and Franklin FA (2001) Secular acceleration of Io derived from mutual satellite events. *Astronomical Journal* 122(5): 2734–2739.
- Albers M and Christensen UR (1996) The excess temperature of plumes rising from the core–mantle boundary. *Geophysical Research Letters* 23: 3567–3570.
- Alley KM and Parmentier EM (1998) The development of mantle convection in an initially stable chemical stratification; Application to a model of lunar mare basalt genesis. *Physics of the Earth and Planetary Interiors* 108: 15–32.
- Anderson DL (1987) A seismic equation of state II: Shear properties and thermodynamics of the lower mantle. *Physics of the Earth and Planetary Interiors* 45: 307–323.
- Anderson JD (1998) Gravitational fields and the interior structure of icy Satellites, *American Physical Society*, Joint Fall Meeting of the Texas Section of the APS, abstract #2P.03.
- Anderson JD, Jurgens RF, Lau EL, Slade MA, III, and Schubert G (1996) Shape and orientation of mercury from radar ranging data. *Icarus* 124(2): 690–697(8).
- Anderson F and Smrekar S (2006) Global mapping of crustal and lithospheric thickness on Venus. *Journal of Geophysical Research* 111(E8) (doi:10.1029/2004JE002395). (issn: 0148-0227).
- Arkani-Hamed J (2004) A coherent model of the crustal magnetic field of Mars. *Journal of Geophysical Research* 109(E9): E09005.
- Arkani-Hamed J and Dymant J (1996) Magnetic potential and magnetization contrasts of Earth's lithosphere. *Journal of Geophysical Research* 101(B5): 11401–11426.
- Arkani-Hamed J, Schaber GG, and Strom RG (1993) Constraints on the thermal evolution of Venus inferred from Magellan data. *Journal of Geophysical Research* 98: 5309–5315.
- Baker J, Bizzarro M, Wittig N, Connelly J, and Haack H (2005) Early planetesimal melting from an age of 4.5662 Gyr for differentiated meteorites. *Nature* 436: 1127.
- Banerdt WB and Golombek MP (2000) Tectonics of the Tharsis Region of Mars: Insights from MGS topography and gravity, *31st Annual Lunar and Planetary Science Conference*, March 13–17, 2000, Houston, TX, abstract no. 2038.
- Basaltic Volcanism Study Project (1981) *Basaltic Volcanism on the Terrestrial Planets*, 1286 pp. New York: Pergamon Press.
- Basilevsky AT, Head JW, Schaber GG, and Strom RG (1997) The resurfacing history of Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1245–1287. Tucson, AZ: University of Arizona Press.
- Berckheimer H, Kampfmann W, Aulbach E, and Schmeling H (1982) Shear modulus and Q of forsterite and dunite near partial melting from forced-oscillation experiments. *Physics of the Earth and Planetary Interiors* 29: 30–41.
- Bercovici D, Schubert G, and Tackley PJ (1993) On the penetration of the 660 km phase change Mantle downflows. *Geophysical Research Letters* 20: 2599–2602.
- Berman DC and Hartmann WK (2002) Recent fluvial, volcanic, and tectonic activity on the Cerberus Plains of Mars. *Icarus* 159: 1–17.
- Binder AB (1986) The binary fission origin of the Moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 499–515. Houston, TX: Lunar and Planetary Institute.
- Binder AB and Lange MA (1980) On the thermal history, thermal state, and related tectonism of a moon of fission origin. *Journal of Geophysical Research* 85: 3194–3208.
- Bleil U and Petersen N (1983) Variations in magnetization intensity and low-temperature titanomagnetite oxidation of ocean floor basalts. *Nature* 301: 384–388.
- Bond TM and Warner MR (2006) Dating Venus: Statistical models of magmatic activity and impact cratering. *37th Annual Lunar and Planetary Science Conference*, 1957.
- Borg LE, Nyquist LE, Wiesmann H, and Shih C-Y (1997) Constraints on Martian differentiation processes from Rb-Sr and Sm-Nd isotopic analyses of the basaltic shergottite QUE94201. *Geochimica et Cosmochimica Acta* 61: 4915–4931.
- Braginsky SI (1964) Magnetohydrodynamics of the Earth core. *Geomagnetism and Aeronomy* 4: 698–712.
- Brandon AD, Walker RJ, Morgan JW, and Goles GG (2000) Re-Os isotopic evidence for early differentiation of the Martian mantle. *Geochimica et Cosmochimica Acta* 64: 4083–4095.
- Breuer D, Hauck S, II, Buske M, Pauer M, Spohn T, and Solomon S (submitted) Interior evolution of Mercury. *Space Science Reviews*.
- Breuer D and Spohn T (2003) Early plate tectonics versus single-plate tectonics: Evidence from the magnetic field history and crust evolution. *Journal of Geophysical Research – Planets* 108(E7): 5072 (doi:10.1029/20002JE001999).
- Breuer D, Spohn T, and Willner U (1993) Mantle differentiation and the crustal dichotomy of Mars. *Planetary and Space Science* 41: 269–283.
- Breuer D and Spohn T (2006) Viscosity of the Martian mantle and its initial temperature: Constraints from crust formation history and the evolution of the magnetic field. *Planetary and Space Science* 54: 153–169.
- Breuer D, Yuen DA, Spohn T, and Zhang S (1998) Three-dimensional models of Martian mantle convection with phase transitions. *Geophysical Research Letters* 25: 229–232.
- Brown CD and Grimm RE (1996) Lithospheric rheology and flexure at Artemis Chasma, Venus. *Journal of Geophysical Research* 101(E5): 12697–12708.
- Bougher SW, Hunten DM and Phillips RJ (eds.) (1997) *Space Science Series: Venus II – Geology, Geophysics, Atmosphere, and Solar Wind Environment*, 1330 pp. Tucson, AZ: The University of Arizona Press.
- Bullock MA, Grinspoon DH, and Head JW (1993) Venus resurfacing rates: Constraints provided by 3-D Monte Carlo simulations. *Geophysical Research Letters* 20: 2147–2150.
- Burov E and Diamant M (1995) The effective elastic thickness (T_e) of the continental lithosphere: What does it really mean? *Journal of Geophysical Research* 100: 3905–3927.
- Buske M (2006) *Dreidimensionale Thermische Evolutionsmodelle für das Innere von Mars und Merkur*. PhD Thesis, Copernicus GmbH (ISBN 3-936586-58-6).
- Busse FH, Richards AA, and Lenardic A (2006) A simple model of high prandtl and high Rayleigh number convection bounded by thin low-viscosity layers. *Geophysical Journal International* 164(1): 160–167.
- Cadek O and Fleitout L (2003) Effect of lateral viscosity variations in the top 300 km on the geoid and dynamic topography. *Geophysical Journal International* 152(3): 566–580 (doi:10.1046/j.1365-246X.2003.01859.x).

- Cameron AGW, Fegley B, Jr., Benz W, and Slattery WL (1988) The strange density of Mercury: Theoretical considerations. In: Vilas F (ed.) *Mercury*, pp. 692–708. Tucson, AZ: University of Arizona Press.
- Campbell BA, Campbell DB, and DeVries CH (1999) Surface processes in the Venus highlands: Results from analysis of Magellan and Arecibo data. *Journal of Geophysical Research* 104(E1): 1897–1916.
- Caristan Y (1982) The transition from high temperature creep to fracture in Maryland diabase. *Journal of Geophysical Research* 87: 6781–6790.
- Cassen P and Reynolds RT (1973) Role of convection in the moon (Solid state convection role in moon from analysis of models with homogeneous initial distribution of radioactive heat sources). *Journal of Geophysical Research* 78: 3203–3215.
- Cassen P, Reynolds RT, Graziani F, Summers AL, McNellis J, and Blalock L (1979) Convection and lunar thermal history. *Physics of the Earth and Planetary Interiors* 19: 183.
- Cassen P, Young RE, Schubert G, and Reynolds RT (1976) Implications of an internal dynamo for the thermal history of Mercury. *Icarus* 28: 501–508.
- Chen JH and Wasserburg GJ (1986) Formation ages and evolution of Shergotty and its parent planet from U-Th-Pb systematics. *Geochimica et Cosmochimica Acta* 50: 955–968.
- Chopelas A and Boehler R (1989) Thermal expansion measurement at very high pressure, systematics, and a case for a chemically homogeneous mantle. *Geophysical Research Letters* 16: 1347.
- Choblet G and Sotin C (2000) 3D thermal convection with variable viscosity: Can transient cooling be described by a quasi-static scaling law? *Physics of the Earth and Planetary Interiors* 119(3–4): 321–336.
- Christensen C (2006) A deep dynamo generating Mercury's magnetic field. *Nature* 444: 1056–1058.
- Cisowski SM and Fuller M (1986) Lunar paleointensities via the IRMs normalization method and the early magnetic history of the moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 411–424. Houston, TX: Lunar and Planetary Science Institute.
- Clifford SM (1993) The hydrologic response of Mars to the onset of a colder climate and to the thermal evolution of its early crust. In: *Workshop on Early Mars: How Warm and How Wet?* Part 1, pp. 7–9.
- Clifford SM and Fanale FP (1985) The thermal conductivity of the Martian crust. *Lunar and Planetary Sciences XVI*, 144–145, abstract.
- Clauzer C and Huenges E (1995) Thermal conductivity of rocks and minerals. In: Ahrens TJ (ed.) *AGU Reference Shelf 3: Rock Physics and Phase Relations – A Handbook of Physical Constants*, pp. 105–126. Washington, DC: American Geophysical Union.
- Collinson DW (1997) Magnetic properties of Martian meteorites: Implications for an ancient Martian magnetic field. *Meteoritics* 803: 32–36.
- Coleman PJ, Jr. and Russell CT (1977) The remanent magnetic field of the moon. *Royal Society of London Philosophical Transactions Series A* 285: 489–506.
- Condie KC (1997) *Plate Tectonics and Crustal Evolution*, 4th edn., 282 pp. Oxford: Butterworth-Heinemann.
- Connerney J, Acuña M, and Ness N (2003) Magnetic mapping of Mars: Year two, *EGS – AGU – EUG Joint Assembly*, Nice, France, 6–11 April 2003, abstract #4445.
- Connerney JEP, Acuña MH, Ness NF, et al. (2005) From the cover: Tectonic implications of Mars crustal magnetism. *Proceedings of the National Academy of Sciences* 102: 14970–14975.
- Connerney JEP, Acuña MH, Wasilewski P, et al. (1999) Magnetic lineations in the ancient crust of Mars. *Science* 284: 794–798.
- Connerney JEP and Ness NF (1988) Mercury's magnetic field and interior. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 494–513. Tucson, AZ: University of Arizona Press.
- Conzelmann V (1999) *Thermische Evolution des Planeten Merkur berechnet unter Anwendung verschiedener Viskositätsgesetze*, PhD Thesis, University of Münster.
- Davaille A and Jaupart C (1993) Transient high-Rayleigh-number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.
- Dombard AJ, Hauck SA, II, Solomon SC, and Phillips RJ (2001) Potential for long-wavelength folding on Mercury, 32nd Annual Lunar and Planetary Science Conference, 12–16 March, 2001, Houston, Texas, abstract no. 2035.
- Donahue TM, Grinspoon DH, Hartle RE, and Hodges RR, Jr. (1997) Ion/neutral escape of hydrogen and deuterium: Evolution of water. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 385–414. Tucson, AZ: University of Arizona Press.
- Dreibus G and Wänke H (1985) Mars: A volatile rich planet. *Meteoritics* 20: 367–382.
- Dunlop DJ and Özdemir Ö (1997) *Rock Magnetism – Fundamentals and Frontiers*. Cambridge, UK: Cambridge University Press.
- Dyal P, Parkin CW, and Daily WD (1974) Magnetism and the interior of the Moon. *Reviews of Geophysics and Space Physics* 12: 568–591.
- Elkins-Tanton LT, Parmentier EM, and Hess PC (2003) Magma ocean fractional crystallization and cumulate overturn in terrestrial planets: Implications for Mars. *Meteoritics and Planetary Science* 38(12): 1753–1771.
- Fegley B (2004) Venus. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, pp. 487–507. Amsterdam: Elsevier.
- Fegley B, Jr. and Cameron AGW (1987) A vaporization model for iron/silicate fractionation in the Mercury protoplanet. *Earth and Planetary Science Letters* 82: 207–222.
- Fegley B and Prinn RG (1989) Estimation of the rate of volcanism on Venus from reaction rate measurements. *Nature* 337: 55–58.
- Fischer HJ and Spohn T (1990) Thermal-orbital histories of viscoelastic models of Io (J1). *Icarus* 83: 39–65.
- Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars Pathfinder. *Science* 278: 1749–1752.
- Forsyth DW and Uyeda S (1975) On the relative importance of the driving forces of plate motions. *Geophysical Journal of the Royal Astronomical Society* 43: 163–200.
- Fowler AC and O'Brien SBG (1996) A mechanism for episodic subduction on Venus. *Journal of Geophysical Research* 101: 4755–4763.
- Frey HV, Roark JH, Shockley KM, Frey EL, and Sakimoto SEH (2002) Ancient lowlands on Mars. *Geophysical Research Letters* 29 (doi:10.1029/2001GL013832).
- Frey H and Schultz RA (1988) Large impact basins and the mega-impact origin for the crustal dichotomy on Mars. *Geophysical Research Letters* 15: 229–232.
- Fuller M and Cisowski S (1987) Lunar paleomagnetism. In: Jacobs J (ed.) *Geomagnetism*, pp. 307–456. San Diego, CA: Academic Press.
- Giampieri G and Balogh A (2002) 'Mercury's thermoelectric dynamo model revisited'. *Planetary and Space Science* 50: 757–762.

- Gosh A and McSween HY (1998) A thermal model for the differentiation of Asteroid 4 Vesta, based on radiogenic heating. *Icarus* 134(2): 187–206.
- Grasset O and Parmentier EM (1998) Thermal convection in a volumetrically heated, infinite Prandtl number fluid with strongly temperature-dependent viscosity: Implications for planetary thermal evolution. *Journal of Geophysical Research* 103: 18171–18181.
- Greeley R and Schneid BD (1991) Magma generation on Mars: Amounts, rates, and comparisons with Earth, Moon and Venus. *Science* 254: 996–999.
- Greeley R and Spudis PD (1981) Volcanism on Mars. *Reviews of Geophysics and Space Physics* 19: 13–41.
- Grimm RE (1994) The deep structure of Venusian plateau highlands. *Icarus* 112(1): 89–103.
- Grosfils EB and Head JW (1996) The timing of giant radiating dike swarms emplacement on Venus: Implications for resurfacing of the planet and its subsequent evolution. *Journal of Geophysical Research* 101: 4645–4656.
- Grossman L (1972) Condensation in the primitive solar nebula. *Geochimica et Cosmochimica Acta* 36: 597–619.
- Grott M, Hauber E, Werner SC, Kronberg P, and Neukum G (2005) High heat flux on ancient Mars: Evidence from Rift Flank Uplift at Coracis Fossae. *Geophysical Research Letters* 32: L21201 (doi:10.1029/2005GL023894).
- Grott M, Hauber E, Werner SC, Kronberg P, and Neukum G (2007) Mechanical modeling of thrust faults in the Thaumasia region, Mars, and implications for the Noachian heat flux. *Icarus* 186: 517–526 (doi:10.1016/j.icarus.2006.10.001).
- Halliday AN, Wänke H, Birck J-L, and Clayton RN (2001) The accretion, composition and early differentiation of Mars. *Space Science Reviews* 96: 197–230.
- Hansen VL and Willis JJ (1996) Structural analysis of a sampling of tesserae: Implications for Venus geodynamics. *Icarus* 23: 296–312.
- Harder H (1998) Phase transitions and the 3D-planform of thermal convection in the Martian mantle. *Journal of Geophysical Research* 103: 16775–16798.
- Harder H (2000) Mantle convection and the dynamic geoid of Mars. *Geophysical Research Letters* 27: 301–304.
- Harder H and Christensen U (1996) A one-plume model of Martian mantle convection. *Nature* 380: 507–509.
- Harder H and Schubert G (2001) Sulfur in Mercury's core? *Icarus* 151: 118–122.
- Hartmann WK, Malin M, McEwen A, et al. (1999) Recent volcanism on Mars from crater counts. *Nature* 397: 586–589.
- Hartmann WK and Neukum W (2001) Cratering chronology and the evolution of Mars. *Space Science Reviews* 96(1/4): 165–194.
- Harmon JK (1997) Mercury radar studies and lunar composition. *Advances in Space Research* 19: 1487–1496.
- Harper CL, Jr., Nyquist LE, Bansal B, Wiesmann H, and Shih CY (1995) Rapid accretion and early differentiation of Mars indicated by 142Nd/144Nd in SNC meteorites. *Science* 267(5195): 213–217.
- Hauck SA, Dombard AJ, Phillips RJ, and Solomon SC (2004) Internal and tectonic evolution of Mercury. *Earth and Planetary Science Letters* 222(3–4): 713–728.
- Hauck SA and Phillips RJ (2002) Thermal and crustal evolution of Mars. *Journal of Geophysical Research* 107(E7) (doi:10.1029/2001JE001801).
- Head JW, Greeley R, Golombek MP, et al. (2001) Geological processes and evolution. *Space Science Reviews* 96(1/4): 263–292.
- Head J and Hiesinger H (1999) Lunar mare stratigraphy and lunar evolution. *Geophysical Research Abstract* 1: 735.
- Heimpel MH, Aurnou JM, Al-Shamali FM, and Gomez Perez N (2005) A numerical study of dynamo action as function of spherical shell geometry. *Earth and Planetary Science Letters* 236: 542–557.
- Helbert J and Benkhoff J (2003) A new approach to assessing the burial depth of ground ice deposits and its application to proposed MER landing sites in Isidis Planitia. *Journal of Geophysical Research* 108(E12): ROV 28–31.
- Helbert J, Mariangeli L, Piccioni G, Drossart P, and Mueller N (2006) The VIRTIS Venus Express Team, observing the surface of Venus with VIRTIS on Venus Express, *American Astronomical Society, DPS meeting #38*, #16.03.
- Herrick RR (1994) Resurfacing history of Venus. *Geology* 22: 703–706.
- Herrick RR and Parmentier EM (1994) Episodic large-scale overturn of two-layer mantle in terrestrial planets. *Journal of Geophysical Research* 99: 2053–2062.
- Herrick RR and Phillips RJ (1994) Implications of a Global Survey of Venusian impact craters. *Icarus* 111: 387–416.
- Hess PC (2000) Petrogenesis of lunar troctolites – Implications for the Moon and its evolution, *31st Annual Lunar and Planetary Science Conference*, 13–17 March, 2000, Houston, Texas, abstract no. 1389.
- Hess PC and Parmentier EM (1995) A model for the thermal and chemical evolution of the Moon's interior: Implications for the onset of mare volcanism. *Earth and Planetary Science Letters* 134(3–4): 501–514.
- Hess PC and Parmentier EM (2001) Implications of magma ocean cumulate overturn for Mars, *32nd Annual Lunar and Planetary Science Conference*, 12–16 March, 2001, Houston, Texas, abstract no. 1319.
- Hide R (1972) Comments on the Moon's magnetism. *Moon* 4: 39.
- Hiesinger H, Jaumann R, Neukum G, and Head JW, III (2000) Ages of mare basalts on the lunar nearside. *Journal of Geophysical Research* 105: 29239–29275.
- Hofmeister AM (1999) Mantle values of thermal conductivity and the geotherm from Phonon lifetimes. *Science* 283(5408): 1699.
- Hood L and Huang Z (1991) Formation of magnetic anomalies antipodal to lunar impact basins: Two dimensional model calculations. *Journal of Geophysical Research* 96: 9837–9846.
- Hood L and Vickery A (1984) Magnetic field amplification and generation in hypervelocity meteoroid impacts with application to lunar paleomagnetism, Proceedings of the Lunar and Planetary Science Conference 15th, Part 1. *Journal of Geophysical Research* 89 supplement: C211–C223.
- Hood LL, Zakharian A, Halekas J, et al. (2001) Initial mapping and interpretation of lunar crustal magnetic anomalies using {Lunar Prospector} magnetometer data. *Journal of Geophysical Research* 106: 27825–27839.
- Hussmann H and Spohn T (2004) Thermal-orbital evolution of Io and Europa. *Icarus* 171(2): 391–410.
- Ito K and Kennedy GC (1971) An experimental study of the basalt-garnet granulite-eclogite transition. In: Heacock JG (ed.) *Geophysical Monograph Series 14: The Structure and Physical Properties of the Earth's Crust*, pp. 303–314. Washington, DC: American Geophysical Union.
- Ivanov MA and Head JW (2006) Testing directional (evolutionary) and non-directional models of the geologic history of Venus: Results of mapping in a geotraverse along the Equator of Venus. *37th Annual Lunar and Planetary Science Conference*, 13–17 March, 2006, League City, TX, abstract no. 1366.
- Jaeger WL, Turtle EP, Keszthelyi LP, Radabaugh J, and McEwen AS (2003) Orogenic tectonism on Io. *Journal of Geophysical Research* 108(E8): 5093.
- Jarvis GT (1984) Time-dependent convection in the Earth's mantle. *Physics of the Earth and Planetary Interiors* 36: 305–327.
- Johnson CL and Richards MA (2003) A conceptual model for the relationship between coronae and large-scale mantle

- dynamics on Venus. *Journal of Geophysical Research (Planets)* 108: 12–21.
- Johnson CL and Sandwell DT (1994) Lithospheric flexure on Venus. *Geophysical Journal International* 119: 627–647.
- Jolliff BL, Gillis JJ, Haskin LA, Korotev RL, and Wieczorek MA (2000) Major lunar crustal terranes: Surface expressions and crust-mantle origins. *Journal of Geophysical Research* 105: 4197–4216.
- Jurdy DM and Stefanick M (2004) Vertical extrapolation of Mars magnetic potentials. *Journal of Geophysical Research* 109(E10): E10005.
- Kaufmann G and Lambeck K (2000) Mantle dynamics, postglacial rebound and the radial viscosity profile. *Physics of the Earth and Planetary Interiors* 121: 301–324.
- Karato S-I, Paterson MS, and Fitzgerald JD (1986) Rheology of synthetic olivine aggregates – Influence of grain size and water. *Journal of Geophysical Research* 91: 8151–8176.
- Karato S and Rubie DC (1997) Towards an experimental study of deep mantle rheology: A new multi-anvil specimen assembly for deformation studies under high pressure and temperatures. *Journal of Geophysical Research* 102: 20111–20122.
- Karato S and Wu P (1993) Rheology of the upper mantle: A synthesis. *Science* 260: 771–778.
- Kaula WM, Schubert G, Lingenfelter RE, Sjogren WL, and Wollenhaupt WR (1972) Analysis and interpretation of lunar laser altimetry. *Proceedings of the Lunar Science Conference* 2: 2189.
- Kiefer WS, Richards MA, Hager BH, and Bills BG (1986) A dynamic model of Venus's gravity field. *Geophysical Research Letters* 13: 14–17.
- Kieffer HH (1976) Soil and surface temperatures at the Viking landing sites. *Science* 194: 1344–1346.
- Kirk RL and Stevenson DJ (1989) The competition between thermal contraction and differentiation in the stress history of the Moon. *Journal of Geophysical Research* 94(B9): 12133–12144.
- Kivelson MG, Khurana KK, Russell CT, et al. (1996) Discovery of Ganymede's magnetic field by the Galileo Spacecraft. *Nature* 384: 537–541.
- Kleine T, Mezger K, Palme H, Scherer E, and Munker C (2004) AGU, Fall Meeting 2004, Abstract P31C-04.
- Kleine T, Münker C, Mezger K, and Palme H (2002) Parid accretion and Early core formation on asteroids and the terrestrial planets from Hf-W chronometry. *Nature* 418: 952–955.
- Konopliv AS and Sjogren WL (1994) Venus spherical harmonic gravity model to degree and order 60. *Icarus* 112(1): 42–54.
- Konrad W and Spohn T (1997) Thermal history of the Moon: Implications for an early core dynamo and post-accretionary magmatism. *Advances in Space Research* 19: 101511–1521(11).
- Kucinskas AB and Turcotte DL (1994) Isostatic compensation of equatorial highlands on Venus. *Icarus* 112(1): 104–116.
- Lavier LL and Steckler MS (1997) The effect of sedimentary cover on the flexural strength of continental lithosphere. *Nature* 389: 476–479.
- Lawrence DJ, Feldman WC, Barraclough BL, et al. (1998) Global element maps of the Moon: The Lunar Prospector gamma-ray spectrometer. *Science* 281: 1484–1489.
- Lee D-C, Halliday AN, Snyder GA, and Taylor LA (1997) Age and origin of the Moon. *Science* 278(5340): 1098.
- Lenardic A, Richards MA, and Busse FH (2006) Depth-dependent rheology and the horizontal length scale of mantle convection. *Journal of Geophysical Research* 111(B7): B07404.
- Lewis JS (1972) The temperature gradient in the solar nebula. *Science* 186: 440–443.
- Liang Y, Price JD, Wark DA, and Watson EB (2001) Nonlinear pressure diffusion in a porous medium: Approximate solutions with applications to permeability measurements using transient pulse decay method. *Journal of Geophysical Research* 106: 529–535.
- Lieske JH (1987) Galilean satellite evolution: Observational evidence for secular changes in mean motions. *Astronomy and Astrophysics* 176: 146–158.
- Lin RP, Mitchell DL, Curtis DW, et al. (1998) Binder lunar surface magnetic fields and their interaction with the solar wind: Results from lunar prospector. *Science* 281: 1480–1484.
- Lister JR and Buffett BA (1995) The strength and efficiency of thermal and compositional convection in the geodynamo. *Physics of the Earth and Planetary Interiors* 91: 17–30.
- Lodders K and Fegley B, Jr. (1998) *The Planetary Scientist's Companion*, pp. 371. Oxford: Oxford University Press.
- Loper DE and Werner C (2000) On the cause of lunar crustal asymmetries, 31st Annual Lunar and Planetary Science Conference, 13–17 March, 2000, Houston, TX, abstract no. 1746.
- Mackwell SJ (1991) High-temperature rheology of enstatite: Implications for creep in the mantle. *Geophysical Research Letters* 18: 2027–2030.
- Matsui T and Abe Y (1986) Origin of the Moon and its early thermal evolution. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 453–468. Houston, TX: Lunar and Planetary Institute.
- McCallum (2001) A new view of the Moon in light of data from Clementine and Prospector missions. *Earth, Moon and Planets* 85–86: 253–269.
- McCammon CA, Ringwood AE, and Jackson I (1983) Thermodynamic of the system Fe-FeO-MgO at high pressure and temperature and a model for the formation of the Earth's core. *Geophysical Journal of the Royal Astronomical Society* 72: 577–595.
- McEwen AS, Keszthelyi LP, Lopes R, Schenk PM, and Spencer JR (2004) The lithosphere and surface of Io. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites and Magnetosphere*, pp. 307–328. Cambridge, UK: Cambridge University Press.
- McEwen AS, Keszthelyi L, Spencer JR, et al. (1998) High-temperature silicate volcanism on Jupiter's moon Io. *Science* 281(5373): 87–90.
- McGill GE and Dimitriou AM (1990) Origin of the Martian global dichotomy by crustal thinning in the late Noachian or early Hesperian. *Journal of Geophysical Research* 95: 12595–12605.
- McGovern PJ, Solomon SC, Smith DE, et al. (2004) Correction to 'Localized gravity/topography admittance and correction spectra on Mars: Implications for regional and global evolution'. *Journal of Geophysical Research* 109(E7): E07007.
- McKenzie D (1977) The initiation of trenches: A finite amplitude instability. In: Talwani M and Piman WC (eds.) *Maurice Ewing Series vol. 1: Island Arcs, Deep Sea Trenches and Back-Arc Basins*, pp. 57–61. Washington, DC: American Geophysical Union.
- McKenzie D (1994) The relationship between topography and gravity on Earth and Venus. *Icarus* 112(1): 55–88.
- McKenzie D, Ford PG, Johnson C, et al. (1992a) Features on Venus generated by plate boundary processes. *Journal of Geophysical Research* 97: 13533–13544.
- McKenzie D, Ford PG, Liu F, and Pettengill GH (1992b) Pancakelike domes on Venus. *Journal of Geophysical Research* 97: 15967–15976.
- McKenzie D and Nimmo F (1997) Elastic thickness estimates for Venus from line of sight accelerations. *Icarus* 130(1): 198–216.
- McKinnon WB, Zahnle K, Ivanov BA, and Melosh HJ (1997) Cratering on Venus: Modeling and observations. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 969–1014. Tucson, AZ: University of Arizona Press.

- McNutt MK (1984) Lithospheric flexure and thermal anomalies. *Journal of Geophysical Research* 89: 11180–11194.
- McSween HY, Jr. (1985) SNC meteorites: Clues to Martian petrologic evolution? *Reviews of Geophysics* 23: 391–416.
- McSween HY, Grove TL, Lentz RCF, et al. (2001) Geochemical evidence for magmatic water within Mars from pyroxenes in Shergotty meteorite. *Nature* 409: 487–490.
- Merk R, Breuer D, and Spohn T (2002) Numerical modeling of ^{26}Al -induced radioactive melting of asteroids considering accretion. *Icarus* 159: 183–191.
- Minar JW and Fletcher CR (1978) Crystallization of a lunar magma ocean. *9th Lunar and Planetary Science Conference*, Houston, TX, 13–17 March, vol. 1, pp. 263–283. New York: Pergamon Press.
- Mitchell DL, Lin RP, Mazelle C, et al. (2001) Probing Mars' crustal magnetic field and ionosphere with the MGS electron reflectometer. *Journal of Geophysical Research* 106(E10): 23419–23428.
- Mitchell DL, Lin RP, Rème H, Acuña MH, Cloutier PA and Ness NF (1999) Crystal magnetospheres observed in the Martian night hemisphere. *American Astronomical Society, DPS Meeting #31*, abstracts #59.04, *Bulletin of the American Astronomical Society*, 31 p. 1584.
- Moore WB (2001) The thermal state of Io. *Icarus* 154: 548–550.
- Moore WB (2003) Tidal heating and convection in Io. *Journal of Geophysical Research* 108: 5096 (doi:10.1029/2002JE001943).
- Moore WB and Schubert G (1997) Venusian crustal and lithospheric properties from nonlinear regression of highland geoid and topography. *Icarus* 128(2): 415–428.
- Moore W, Schubert G, Anderson JD, and Spencer JR (2006) The interior of Io. In: Lopes RMC and Spencer JR (eds.) *Io after Galileo: A New View of Jupiter's Volcanic Moon*, pp. 89–108. Chichester, UK: Springer and Praxis Publication.
- Moresi L-N and Solomatov VS (1995) Numerical investigation of 2D convection with extremely large viscosity variations. *Physics of Fluids* 7: 2154–2162.
- Moresi L-N and Solomatov VS (1998) Mantle convection with a brittle lithosphere: Thoughts on the global tectonic style of the Earth and Venus. *Geophysical Journal* 133: 669–682.
- Morris S and Canright D (1984) A boundary-layer analysis of Benard convection in a fluid of strongly temperature-dependent viscosity. *Physics of the Earth and Planetary Interiors* 36: 355–373.
- Mouginis-Mark PJ, Wilson L, and Zuber MT (1992) The physical volcanology of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 424–452. Tucson, AZ: University of Arizona Press.
- Murakami M, Hirose K, Sata N, Ohishi Y, and Kawamura K (2004) Phase transition of MgSiO₃ perovskite in the deep lower mantle. *Science* 303: 855–858.
- Murthy VR, van Westrenen W, and Fei Y (2003) Experimental evidence that potassium is a substantial radioactive heat source in planetary cores. *Nature* 423: 163–165.
- Namiki N and Solomon SC (1994) Impact crater densities on volcanoes and coronae on Venus: Implications for volcanic resurfacing. *Science* 265: 929–933.
- Neukum G, Jaumann R, Hoffmann H, et al. (2004) Recent and episodic volcanic and glacial activity on Mars revealed by the high resolution Stereo Camera. *Nature* 432: 971–979 (doi:10.1038/nature03231).
- Neumann GA, Zuber MT, Wieczorek MA, McGovern PJ, Lemoine FG, and Smith DE (2004) Crustal structure of Mars from gravity and topography. *Journal of Geophysical Research* 109: E08002 (doi:10.1029/2004JE002262).
- Ness NF, Acuña MH, Connerney J, et al. (1999) MGS magnetic fields and electron reflectometer investigation: Discovery of paleomagnetic fields due to crustal remanence. *Advances in Space Research* 23(11): 1879–1886.
- Ness NF, Behannon KW, Lepping RP, Schatten KH, and Whang YC (1974) Magnetic field observations near Mercury: Preliminary results from Mariner 10. *Science* 185: 151–160.
- Ness NF, Behannon KW, Lepping RP, and Whang YC (1975) Magnetic field of Mercury. *Journal of Geophysical Research* 80: 2708–2716. (Part I).
- Ness NF, Behannon KW, Lepping RP, and Whang YC (1976) Observations of Mercury's magnetic field. *Icarus* 28: 479–488.
- Nimmo F (2000) Dike intrusion as a possible cause of linear Martian magnetic anomalies. *Geology* 28: 391–394.
- Nimmo F (2002) Crustal analysis of Venus from Magellan satellite observations at Atalanta Planitia, Beta Regio, and Thetis Regio. *Geology* 30: 987–990.
- Nimmo F and McKenzie D (1996) Modelling plume-related uplift, gravity and melting on Venus. *Earth and Planetary Science Letters* 145(1–4): 109–123.
- Nimmo F and McKenzie D (1998) Volcanism and tectonism on Venus. *Annual Review of Earth and Planetary Sciences* 26: 23–53.
- Nimmo F and Stevenson D (2000) Influence of early plate tectonics on the thermal evolution and magnetic field of Mars. *Journal of Geophysical Research* 105: 11969–11979.
- Nimmo F and Stevenson D (2001) Estimates of Martian crustal thickness from viscous relaxation of topography. *Journal of Geophysical Research* 106: 5085–5098.
- Nimmo F and Tanaka K (2005) Early crustal evolution of Mars. *Annual Review of the Earth and Planetary Sciences* 33: 133–161.
- Nimmo F and Watters TR (2004) Depth of faulting on Mercury: Implications for heat flux and crustal and effective elastic thickness. *Geophysical Research Letters* 31: L02701 (doi:02710.01029/02003GL018847).
- Norman MD (2002) Thickness and composition of the Martian crust revisited: Implications of an ultradepleted mantle with a Nd isotopic composition like that of QUE94201, 33rd Annual Lunar and Planetary Science Conference, 11–15 March, 2002, Houston, TX, abstract no. 1175.
- Nozette S, Rustan P, Pleasance LP, et al. (1994) The Clementine mission to the Moon: Scientific overview. *Science* 266: 1835–1839.
- Oganov AR and Ono S (2004) Theoretical and experimental evidence for a post-perovskite phase of MgSiO₃ in Earth's D'' layer. *Nature* 430: 445–448.
- Ogawa M (2000) Numerical models of magmatism in convecting mantle with temperature-dependent viscosity and their implications for Venus and Earth. *Journal of Geophysical Research* 105: 6997–7012.
- O'Reilly TC and Davies GF (1981) Magma transport of heat on Io: A mechanism allowing a thick lithosphere. *Geophysical Research Letters* 8: 313–316.
- Palme H, Spettel B, Winke H, Bischoff A, and StGfler D (1984) Early differentiation of the Moon: Evidence from trace elements in Plagioclase, *Proceedings of the Lunar Planetary Science Conference*, 15th, Houston, TX, 12–16 March, 1984 (*Journal of Geophysical Research* 89: C3–C15).
- Pavri B, Head JW, Klose KB, and Wilson L (1992) Steep-sided domes on Venus: Characteristics, geologic setting and eruption conditions from Magellan data. *Journal of Geophysical Research* 97: 13445–13478.
- Parmentier EM and Hess PC (1992) Chemical differentiation of a convecting planetary interior: Consequences for a one plate planet such as Venus. *Geophysical Research Letters* 19: 2015–2018.
- Parmentier EM, Zhong S, and Hess PC (1999) Asymmetric evolution of the Moon: A possible consequence of chemical differentiation, *Workshop on New Views of the Moon II: Understanding the Moon Through the Integration of Diverse*

- Datasets, 22–24 September, 1999, Flagstaff, AZ, abstract no. 8034.
- Parmentier EM, Zhong S, and Zuber M (2002) Gravitational differentiation due to initial chemical stratification: Origin of lunar asymmetry by the creep of dense KREEP? *Earth and Planetary Science Letters* 201: 473–480.
- Pauer M, Fleming K, and Cadek O (2006) Modeling the dynamic component of the geoid and topography of Venus. *Journal of Geophysical Research* (doi:10.1029/2005JE002511).
- Peale SJ, Cassen P, and Reynolds RT (1979) Melting of Io by tidal dissipation. *Science* 203: 892–894.
- Peltier WR and Jiang X (1996) Glacial isostatic adjustment and Earth rotation: Refined constraints on the viscosity of the deepest mantle. *Journal of Geophysical Research* 101(B2): 3269–3290.
- Phillips RJ (1994) Estimating lithospheric properties at Atlas Regio, Venus. *Icarus* 112(1): 147–170.
- Phillips RJ, Bullock MA, and Hauck SA (2001) Climate and interior coupled evolution on Venus. *Geophysical Research Letters* 28: 1779–1782.
- Phillips RJ and Hansen VL (1998) Geological evolution of Venus: Rises, plains, plumes, and plateaus. *Science* 279: 1492–1497.
- Phillips RJ, Raubertas RF, Arvidson RE, et al. (1992) Impact craters and Venus resurfacing history. *Journal of Geophysical Research* 97: 15923–15948.
- Phillips RJ, Zuber MT, Solomon SC, et al. (2001) Ancient geodynamics and global-scale hydrology on Mars. *Science* 291(5513): 2587–2591.
- Pritchard ME and Stevenson DJ (1999) Thermal aspects of a lunar origin by giant impact. In: Canup R and Righter K (eds.) *The Origin of the Earth and Moon*, pp. 179–196. Tucson, AZ: University of Arizona Press.
- Purucker M, Ravat D, Frey H, Voorhies C, Sabaka T, and Acuña M (2000) An altitude-normalized magnetic map of Mars and its interpretation. *Geophysical Research Letters* 27: 2449–2452.
- Ratcliff JT, Tackley PJ, Schubert G, and Zebib A (1997) Transitions in thermal convection with strongly variable viscosity. *Physics of the Earth and Planetary Interiors* 102: 201–212.
- Reese CC, Solomatov VS, and Moresi L-N (1999) Non-Newtonian stagnant lid convection and magma resurfacing on Venus. *Icarus* 139: 67–80.
- Reese CC, Solomatov VS, and Moresi L-N (1998) Heat transport efficiency for stagnant lid convection with dislocation viscosity: Application to Venus and Mars. *Journal of Geophysical Research* 103: 13643–13658.
- Richter FM (1978) Experiments on the stability of convection rolls in fluids whose viscosity depends on temperature. *Journal of Fluid Mechanics* 89: 553–560.
- Ringwood AE (1975) Some aspects of the minor element chemistry of lunar mare basalts. *The Moon* 12: 127–157.
- Ringwood AE (1977) Composition of the core and implications for the origin of the Earth. *Geochemical Journal* 11: 111–135.
- Roberts GO (1979) Fast viscous Bernard convection. *Geophysical and Astrophysical Fluid Dynamics* 12: 235–272.
- Roberts JH and Zhong SJ (2004) Degree-1 mantle convection as a process for generating the Martian hemispheric dichotomy. In: *Workshop on Hemispheres Apart: The origin and Modification of the Martian Crustal Dichotomy*, LPI Contribution No. 1213, 52–53, 2004.
- Roberts JH and Zhong SJ (2006) Degree-1 convection in the Martian mantle and the origin of the hemisphere dichotomy. *Journal of Geophysical Research* 111(E6): E06013.
- Rochette P, Lorand J-P, Fillion G, and Sautter V (2001) Pyrrhotite and the remanent magnetization of SNC meteorites: A changing perspective on Martian magnetism. *Earth and Planetary Science Letters* 190(1–2): 1–12.
- Runcorn SK (1975) An ancient lunar magnetic dipole field. *Nature* 253: 701–703.
- Runcorn SK (1977) Early melting of the Moon. In *Proceedings of the Lunar Science Conference 8th*, Houston, TX, 14–18 March, 1977, pp. 463–467.
- Russell CT, Elphic RC, and Slavin JA (1979a) Initial Pioneer Venus magnetic field results: Dayside observations. *Science* 203: 745.
- Russell CT, Elphic RC, and Slavin JA (1979b) Initial Pioneer Venus magnetic field results: Nightside observations. *Science* 205: 114.
- Russell CT, Baker DN, and Slavin JA (1988) The magnetosphere of Mercury. In: Violas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 514–561. Tucson, AZ: University Press of Arizona.
- Sanloup C, Jambon A, and Gillet P (1999) A simple chondritic model of Mars. *Physics of the Earth and Planetary Interiors* 112: 43–54.
- Sandwell DT and Schubert G (1992) Evidence for retrograde lithospheric subduction on Venus. *Science* 257(5071): 766–770.
- Schaber GG, Strom RG, Moore HJ, et al. (1992) Geology and distribution of impact craters on Venus: What are they telling us? *Journal of Geophysical Research* 97: 13257–13301.
- Schenk P, Hargitai H, Wilson R, McEwen A, and Thomas P (2001) The mountains of Io: Global and geological perspectives from Voyager and Galileo. *Journal of Geophysical Research* 106: 33201–33222.
- Schott B, van den Berg AP, and Yuen DA (2001) Focussed time-dependent Martian volcanism from chemical differentiation coupled with variable thermal conductivity. *Geophysical Research Letters* 28: 224271–224274.
- Schubert G (1979) Subsolidus convection in the mantles of terrestrial planets. *Annual Review of the Earth and Planetary Sciences* 7: 289–342.
- Schubert G and Anderson CA (1985) Finite element calculations of very high Rayleigh number thermal convection. *Geophysical Journal of the Royal Astronomical Society* 80: 575–601.
- Schubert G, Bercovici D, and Glatzmeier GA (1990) Mantle dynamics in Mars and Venus: Influence of an immobile lithosphere on three-dimensional mantle convection. *Journal of Geophysical Research* 95: 14105–14129.
- Schubert G, Cassen P, and Young RE (1979a) Core cooling by subsolidus mantle convection. *Physics of the Earth and Planetary Interiors* 20: 194–208.
- Schubert G, Cassen P, and Young RE (1979b) Subsolidus convective cooling histories of terrestrial planets. *Icarus* 38: 192–211.
- Schubert G and Lingenfelter RE (1973) Martian center of mass – Center of figure offset. *Nature* 242: 251–252.
- Schubert G, Ross MN, Stevenson DJ, and Spohn T (1988) Mercury's thermal history and the generation of its magnetic field. In: Violas F, Chapman CR, and Matthews MS (eds.) *Mercury*, pp. 514–561. Tucson, AZ: University Press of Arizona.
- Schubert G, Russell CT, and Moore WB (2000) Timing of the Martian dynamo. *Nature* 408: 666–667.
- Schubert G, Solomatov VS, Tackley PJ, and Turcotte DL (1997) Mantle convection and the thermal evolution of Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1245–1287. Tucson, AZ: University of Arizona Press.
- Schubert G, Solomon SC, Turcotte DL, Drake MJ, and Sleep NH (1992) Origin and thermal evolution of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 147–183. Tucson, AZ: University of Arizona Press.
- Schubert G and Spohn T (1990) Thermal history of Mars and the sulfur content of its core. *Journal of Geophysical Research* 95: 14095–14104.

- Schubert G, Spohn T, and Reynolds RT (1986) Thermal histories, compositions, and internal structures of the moons of the solar system. In: Burns JA and Matthews M (eds.) *Satellites*, pp. 224–292. Tucson, AZ: University of Arizona Press.
- Schubert G, Turcotte DL, and Olson P (2001) *Mantle Convection in the Earth and Planets*, 940 pp. Cambridge, UK: Cambridge University Press.
- Schubert G, Turcotte DL, and Oxburgh ER (1970) Phase change instability in the mantle. *Science* 169: 1075–1077.
- Schumacher S and Breuer D (2006) Influence of a variable thermal conductivity on the thermochemical evolution of Mars. *Journal of Geophysical Research* 111(E2): E02006.
- Schutt DL and Lesher CE (2006) Effects of melt depletion on the density and seismic velocity of garnet and spinel lherzolite. *Journal of Geophysical Research* 111(B5): B05401.
- Scott DR and Stevenson DJ (1986) Magma ascent by porous flow. *Journal of Geophysical Research* 91: 9283–9296.
- Segatz M, Spohn T, Ross MN, and Schubert G (1988) Tidal dissipation, surface heat flow, and figure of viscoelastic models of Io. *Icarus* 75: 187–206.
- Seipold U (1992) Depth dependence of thermal transport properties for typical crustal rocks. *Physics of the Earth and Planetary Interiors* 69(3–4): 299–303.
- Seipold U (1998) Temperature dependence of thermal transport properties of crystalline rocks – a general law. *Tectonophysics* 291: 161–171.
- Senshu H, Kuramoto K, and Matsui T (2002) Thermal evolution of a growing Mars. *Journal of Geophysical Research* 107: 5118 (doi:10.1029/2001JE001819).
- Sharpe HN and Peltier WR (1978) Parameterized mantle convection and the Earth's thermal history. *Geophysical Research Letters* 5: 737–740.
- Sharpe HN and Strangway DW (1976) The magnetic field of Mercury and modes of thermal evolution. *Geophysical Research Letters* 3: 285–288.
- Siegfried RW and Solomon SC (1974) Mercury – Internal structure and thermal evolution. *Icarus* 23: 192–205.
- Simons M, Hager BH, and Solomon SC (1994) Global variations in the geoid:topography admittance of Venus. *Science* 264: 798–803.
- Simons M, Solomon SC, and Hager BH (1997) Localization of gravity and topography: Constraints on the tectonics and mantle dynamics of Venus. *Geophysical Journal International* 131: 24–31.
- Sleep NH (1994) Martian plate tectonics. *Journal of Geophysical Research* 99: 5639–5655.
- Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implication for surface evolution. *Science* 284: 1495–1503.
- Smrekar SE (1994) Tectonics of mantle plumes on Venus. *Lunar and Planetary Institute*, abstract, pp. 72.
- Smrekar SE and Stofan ER (1999) Origin of corona-dominated topographic rises on Venus. *Icarus* 139: 100–115.
- Sohl F and Spohn T (1997) The structure of Mars: Implications from SNC-Meteorites. *Journal of Geophysical Research* 102: 1613–1635.
- Sohl F, Schubert G, and Spohn T (2005) Geophysical constraints on the composition and structure of the Martian interior. *Journal of Geophysical Research* 110(E12): E12008.
- Solomatov VS (1995) Scaling of temperature- and stress-dependent viscosity. *Physics of Fluids* 7: 266–274.
- Solomatov VS and Moresi L-N (1996) Stagnant lid convection on Venus. *Journal of Geophysical Research* 101: 4737–4753.
- Solomon SC (1976) Some aspects of core formation in Mercury. *Icarus* 28: 509–521.
- Solomon SC (1977) The relationship between crustal tectonics and internal evolution in the Moon and Mercury. *Physics of the Earth and Planetary Interiors* 15: 135–145.
- Solomon SC (1979) Formation, history and energetics of core in terrestrial planets. *Physics of the Earth and Planetary Interiors* 19: 168–182.
- Solomon CS, Aharonson O, Aurnou JM, et al. (2005) New perspectives on ancient Mars. *Science* 307: 1214–1220.
- Solomon SC and Longhi J (1977) Magma oceanography. Part 1: Thermal evolution. *Proceedings of the Lunar Science Conference 8th*, Houston, TX, 583–599.
- Spohn T (1991) Mantle differentiation and thermal evolution of Mars, Mercury, and Venus. *Icarus* 90: 222–236.
- Spohn T, Acuña MA, Breuer D, et al. (2001) Geophysical constraints on the evolution of Mars. *Space Science Reviews* 96: 231–262.
- Spohn T and Breuer D (1993) Mantle differentiation through continental crust growth and recycling and the thermal evolution of the Earth. In: Takahashi E, Jeanloz R, and Rubie DC (eds.) *Geophysical Monograph Series 74, IUGG vol. 14: Evolution of the Earth and Planets*, pp. 55–71. Washington DC: American Geophysical Union.
- Spohn T, Sohl F, and Breuer D (1998) Mars. *Astronomy and Astrophysics Review* 8: 181–235.
- Spohn T, Sohl F, Wieczorkowski K, and Conzelmann V (2001) The interior structure of Mercury: What we know, what we expect from BepiColombo. *Planetary Space Science* 49: 1561–1570.
- Sprague AL, Nash DB, Witteborn RC, and Cruikshank DP (1997) Mercury's feldspar connection: Mid-IR measurements suggest plagioclase. *Advances in Space Research* 19(10): 1507–1510.
- Stanley S, Bloxham J, Hutchinson WE, and Zuber MT (2005) Thin shell dynamo models consistent with Mercury's weak observed magnetic field. *Earth and Planetary Science Letters* 234: 27–38.
- Stegman DR, Jellinek AM, Zatman SA, Baumgardner JR, and Richards MA (2003) An early lunar core dynamo driven by thermochemical mantle convection. *Nature* 421(6919): 143–146.
- Steinbach V and Yuen DA (1992) The effects of multiple phase transitions on Venusian mantle convection. *Geophysical Research Letters* 19: 2243–2246.
- Stephenson A, Runcorn SK, and Collinson DW (1975) On changes in the intensity of the ancient lunar magnetic field. *Proceedings of the 6th Lunar Science Conference, Geochimica et Cosmochimica Acta*.
- Stevenson DJ (1987) Origin of the Moon-The collision hypothesis. *Annual Reviews of Earth and Planetary Science* 15: 271–315.
- Stevenson DJ (1990) Fluid dynamics of core formation. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 231–249. New York: Oxford University Press.
- Stevenson DJ (2000) Core superheat. *EOS Transactions of the American Geophysical Union* 81(48): 2000.
- Stevenson DJ (2001) Mars core and magnetism. *Nature* 412: 214–219.
- Stevenson DJ (2002) Possible Connections between the History of the Venus Magnetic Field and Observable Features, American Geophysical Union, Spring Meeting 2002, abstract #P21A-08.
- Stevenson DJ, Spohn T, and Schubert G (1983) Magnetism and thermal evolution of the terrestrial planets. *Icarus* 54: 466–489.
- Stofan ER, Bindschadler DL, Head JW, and Parmentier EM (1991) Corona structures on venus – Models of origin. *Journal of Geophysical Research* 96(E4): 20933–20946.
- Strom RG, Trask NJ, and Guest JE (1975) Tectonism and volcanism on Mercury. *Journal of Geophysical Research* 80: 2478–2507.
- Tackley PJ (1996) On the ability of phase transitions and viscosity layering to induce long wavelength heterogeneity in the mantle. *Geophysical Research Letters* 23: 1985–1988.

- Tackley PJ, Stevenson DJ, Glatzmeier GA, and Schubert G (1993) Effects of an endothermic phase transition at 670 km depth in a spherical model of convection in the Earth's mantle. *Nature* 361: 699–704.
- Tackley PJ, Stevenson DJ, Glatzmeier GA, and Schubert G (1994) Effects of multiple phase transitions in a three-dimensional spherical model of convection in Earth's mantle. *Journal of Geophysical Research* 99: 15877–15901.
- Tanaka LK, Scott DH, and Greeley R (1992) Global stratigraphy. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 345–382. Tucson, AZ: University of Arizona Press.
- Taylor LA, Shervais JW, Hunter RH, Shih D-Y, Bansal BM, and Nyquist LE (1983) Pre 4.2-AE mare basalt volcanism in the lunar highlands. *Earth and Planetary Science Letters* 66: 33–47.
- Toft PB and Arkani-Hamed J (1992) Magnetization of the Pacific Ocean lithosphere deduced from Magsat data. *Journal of Geophysical Research* 97: 4387–4406.
- Toksöz MN, Hsui AT, and Johnston DH (1978) Thermal evolution of the Moon and the terrestrial planets. In: *The Soviet-American Conference on Cosmochemistry of the Moon and Planets*, NASA SP-370, 245–328.
- Tonks WB and Melosh HJ (1990) The physics of crystal settling and suspension in a turbulent magma ocean. In: Newsom HE and Jones JH (eds.) *Origin of the Earth*, pp. 151–174. New York: Oxford University Press.
- Tonks WB and Melosh HJ (1992) Core formation by giant impacts. *Icarus* 100: 326–346.
- Tozer DC (1965) Heat transfer and convection currents. In: *A symposium on continental drift. Philosophical Transactions of the Royal Society* 258: 252–271.
- Tsuchiya T, Tsuchiya J, Umemoto K, and Wentzcovitch RM (2004) Phase transition in MgSiO₃ perovskite in the Earth's lower mantle. *Earth and Planetary Science Letters* 224: 241–248.
- Turcotte DL (1982) Thermal Stresses in Planetary Elastic Lithospheres. *Lunar Planet Science XII*, 809–810, abstract.
- Turcotte DL (1989) Crustal deformation: Earth VS Venus. In: *Lunar and Planetary Institute*, abstracts for the Venus Geoscience Tutorial and Venus Geologic Mapping Workshop, pp. 49.
- Turcotte DL (1993) An episodic hypothesis for Venusian tectonics. *Journal of Geophysical Research* 98: 17061–17068.
- Turcotte DL and Huang J (1990) Implications on crust formation on Mars from parameterized convection calculations. *Lunar and Planetary Science Conference Abstract* 21: 1266.
- Turcotte DL and Oxburgh ER (1967) Finite amplitude convective cells and continental drift. *Journal of Fluid Mechanics* 28: 29–42.
- Turcotte DL and Schubert G (1982) *Geodynamics*. London: Wiley.
- van Thienen P, Vlaar NJ, and van den Berg AP (2005) Assessment of the cooling capacity of plate tectonics and flood volcanism in the evolution of Earth, Mars and Venus. *Physics of the Earth and Planetary Interiors* 150: 287–315.
- Veeder GJ, Matson DL, Johnson TV, Davies AG, and Blaney DL (2004) The polar contribution to the heat flow of Io. *Icarus* 169(1): 264–270.
- Vilas F (1988) Surface composition of Mercury from reflectance spectrometry. In: Vilas F, Chapman CR, and Matthews MS (eds.) *Mars*, pp. 622–650. Tucson, AZ: University of Arizona Press.
- Wark DA and Watson EB (1998) Grain-scale permeabilities of texturally equilibrated, monomineralic rocks. *Earth and Planetary Science Letters* 164: 591–605.
- Warren PH (1985) The magma ocean concept and lunar evolution. *Annual Review of the Earth and Planetary Sciences* 13: 201–240.
- Watters TR, Robinson MS, Bina CR, and Spudis PD (2004) Thrust faults and the global contraction of Mercury. *Geophysical Research Letters* 31: 04701.
- Warren PH and Wasson JT (1979) The origin on KREEP. *Reviews of Geophysics* 17: 73–88.
- Watts AB (2001) *Isostasy and Flexure of the Lithosphere*. New York: Cambridge University Press.
- Weertman J and Weertman JR (1975) High temperature creep of rock and mantle viscosity. *Annual Review of the Earth and Planetary Sciences* 3: 293–315.
- Weidenschilling SJ (1978) Iron/silicate fractionation and the origin of Mercury. *Icarus* 35: 99–111.
- Weinstein SA (1995) The effects of a deep mantle endothermic phase change on the structure of thermal convection in silicate planets. *Journal of Geophysical Research* 100(E6): 11719–11728.
- Weinstein SA (1996) The potential role of non-Newtonian rheology in the resurfacing of Venus. *Geophysical Research Letters* 23: 511–514.
- Weiss BP, Vali H, Baudenbacher FJ, Kirschvink JL, Stewart ST, and Shuster DL (2002) Records of an ancient Martian magnetic field in ALH84001. *Earth and Planetary Science Letters* 201(3–4): 449–463.
- Weizman A, Stevenson DJ, Prialnik D, and Podolak M (2001) Modeling the volcanism on Mars. *Icarus* 150: 195–205.
- Wenzel MJ, Manga M, and Jellinek AM (2004) Tharsis as a consequence of Mars' dichotomy and layered mantle. *Geophysical Research Letters* 31(4): L04702.
- Wetherill GW (1985) Occurrence of giant impacts during the growth of the terrestrial planets. *Science* 228: 877–879.
- Wetherill GW (1988) Accumulation of Mercury from planetesimals. In: Vilas F (ed.) *Mercury*, pp. 670–691. Tucson, AZ: University of Arizona Press.
- Wieczorek MA and Phillips RJ (2000) The 'Procellarum KREEP Terrane': Implications for mare volcanism and lunar evolution. *Journal of Geophysical Research* 105: 20417–20430.
- Williams J-P and Nimmo F (2004) Thermal evolution of the Martian core: Implications for an early dynamo. *Geology* 32: 97–100.
- Wise DU, Golombek MP, and McGill GE (1979) Tectonic evolution of Mars. *Journal of Geophysical Research* 84: 7934–7939.
- Wilhelms DE and Squyres SW (1984) The Martian hemispheric dichotomy may be due to a giant impact. *Nature* 309: 138–140.
- Wood JA (1970) Nature and Interpretation of the Apollo 11 Lunar Samples. *Physics of the Solar System* 353.
- Wood JA, Dickey JS, Marvin UB and Powell BN (1970) Lunar anorthosites and a geophysical model of the Moon, *Proceedings of the Apollo 11 Lunar Science Conference*, 965–988.
- Yoder CF (1979) How tidal heating in Io drives the Galilean orbital resonance locks. *Nature* 279: 767–770.
- Yoder CF and Peale SJ (1981) The tides of Io. *Icarus* 47: 1–35.
- Zebib A, Goyal AK, and Schubert G (1985) Convective motion in a spherical shell. *Journal of Fluid Mechanics* 152: 39–48.
- Zebib A, Schubert G, Dein JL, and Paliwal RC (1983) Character and stability of axisymmetric thermal convection in spheres and spherical shells. *Geophysical and Astrophysical Fluid Dynamics* 23: 1–42.
- Zhong S and Zuber MT (2001) Degree-1 mantle convection and the crustal dichotomy on Mars. *Earth and Planetary Science Letters* 189: 75–84.
- Zuber MT, Solomon SC, Phillips RJ, et al. (2000) Internal structure and early thermal evolution of Mars from Mars Global Surveyor topography and gravity. *Science* 287: 1788–1793.

10.10 Solid Planet–Atmosphere Interactions

M. Yu. Zolotov, Arizona State University, Tempe, AZ, USA

© 2007 Elsevier B.V. All rights reserved.

10.10.1	Atmosphere–Surface Interplay on Solar System Bodies	349
10.10.2	Observational Constraints on Venus and Mars	350
10.10.2.1	Physical and Chemical Signs of Surface Rock Alteration	350
10.10.2.2	Atmospheric Compositions in the Context of Atmosphere–Surface Interactions	353
10.10.3	Chemistry of Atmosphere–Surface Reactions on Mars and Venus	355
10.10.3.1	Redox Processes	355
10.10.3.1.1	Oxidation on Venus	355
10.10.3.1.2	Oxidation on Mars	356
10.10.3.2	Reactions with Sulfur Gases	357
10.10.3.3	Reactions with H ₂ O	359
10.10.3.4	Reactions with CO ₂	359
10.10.3.5	Reactions with Halogens	360
10.10.3.6	Condensation of Metal and Nonmetal Species, the Role of Altitude	361
10.10.3.7	Kinetic Aspects of Surface–Gas-Type Reactions	361
10.10.3.8	Local Reactions Caused by Impacts and Volcanic Eruptions	362
10.10.4	Wind-Related Processes	363
10.10.5	Atmosphere–Surface Interactions throughout History	363
10.10.6	Summary and Unsolved Questions	364
References		365

10.10.1 Atmosphere–Surface Interplay on Solar System Bodies

Even without an abundant aqueous phase, planetary surfaces are coupled with planetary atmospheres through complex physical and chemical processes. Volcanic degassing, impacts, and seepage from mantle and crustal rocks supply gases into atmospheres. Condensation of released volatiles affects abundances of atmospheric gases and surface composition. Compositions of atmospheres and surfaces are also influenced by chemical interactions of gases with surface rocks and liquids, if they are present. In some cases, atmospheric abundances of volatiles can be controlled (buffered) by condensation/sublimation equilibria or chemical reactions at the surface. Rates of chemical reactions at the atmosphere–surface interface are affected by temperatures that, in turn, are influenced by solar luminosity, cloudiness, and the greenhouse effect. In dense atmospheres, eolian processes contribute to weathering, mass transfer, and separation of surface solids. Throughout history, atmosphere–surface interactions could have been affected by changing solar

luminosity, orbital parameters, internal processes, impact cratering, geological resurfacing, as well as gas escape into space.

Three major terrestrial planets – Venus, Earth, and Mars – reveal a broad-scale coupling of atmospheric and solid-body processes. On Earth, the major pathways of endogenic and exogenic processes are mainly determined by mantle convection and plate tectonics (*see Chapter 6.02*), the hydrosphere, life, and impacts. On Mars and Venus, surface composition and climatic evolution are affected by volcanism, impacts, and atmosphere–surface interactions with very limited and/or temporal influences from aqueous processes. This chapter is mainly devoted to atmosphere–surface interactions on Venus and Mars.

Despite significant differences in surface temperature, pressure, and atmospheric composition (**Table 1**), the atmosphere–surface interactions on Venus and Mars have several common characteristics. Aqueous solutions are unstable on both of their surfaces. On both planets, surfaces have not been significantly affected by volcanic, impact, and eolian processes for hundreds of millions of years. Nevertheless, surface materials reveal signs of alteration by coupled physical

Table 1 Environmental conditions and the composition of major and chemically active gases in lower atmospheres of Venus and Mars

	Venus ^a	Mars ^b
Temperature	740 K	210 K
Pressure	95.6 bar	5.6×10^{-3} bar
CO ₂	96.5 ± 0.8%	95.3%
N ₂	3.5 ± 0.8	2.7 ± 0.5%
⁴⁰ Ar	31^{+20}_{-10} ppm	$1.6^{+0.3}_{-0.2}$ ppm
³⁶ + ³⁸ Ar	35.5^{+20}_{-10} ppm	5.3 ± 1 ppm
O ₂	c	0.13 ± 0.01%
SO ₂	150 ± 30 ppm (22–42 km)	<0.1 ppm
H ₂ O	30 ± 15 ppm (0–45 km)	0.03% ^d
CO	17 ± 1 ppm (17 km)	0.07 ± 0.014%
COS	4.4 ± 1 ppm (33 km)	<0.01 ppm
H ₂ S	3 ± 2 ppm (<20 km)	<0.1 ppm
HCl	0.5 ppm (35–45 km)	<0.1 ppm
S _{1–8}	20 ppb (<50 km)	
HF	4.5 ppb (35–45 km)	
H ₂		15 ± 5 ppm
CH ₄		10 ± 3 ppb
H ₂ O ₂		40 ± 5 ppb
O ₃		18 ± 0.4 ppb
D/H	0.016 ± 0.002 ^e 0.019 ± 0.006	(7.8 ± 0.3) × 10 ⁻⁴

^aVenus temperature and pressure correspond to the modal planetary radius (6051.4 km). Gas abundances represent values closer to the surface. Concentrations of CO and COS are altitude dependent. Near-surface concentrations of these and other gases may be different.

^bMars surface temperature and pressure are estimates of the global annual averages (Kieffer *et al.*, 1992).

^cNear the surface, O₂ pressure is evaluated as $\sim 10^{-21}$ bar (Zolotov, 1996; Fegley *et al.*, 1997b).

^dVariable with season and location.

^eD/H ratios for Venus represent Pioneer Venus *in situ* measurements and ground spectroscopy, respectively.
ppm, parts per million by volume; ppb, parts per billion by volume.

Sources: Esposito *et al.* (1997), Fegley (2003), McSween (2003), Owen (1992), Clancy *et al.* (2004), Encrenaz *et al.* (2004), Krasnopolsky and Feldman (2001), and Krasnopolsky *et al.* (2004).

and chemical processes. Chemical alteration could have included oxidation of primary ferrous (Fe^{2+} -bearing) minerals and sulfides, consumption of sulfur, halogens, and water from the atmosphere, as well as condensation of species with different volatilities. Physical processes include wind abrasion, generation of fine-grained surface materials, mechanical separation of mineral grains, and formation of eolian features, including layered deposits with low porosity. Atmosphere–surface interactions are altitude and latitude dependent, owing to variations in surface temperature, near-surface pressure, and concentrations of atmospheric gases. On both planets, the composition of atmosphere and surface materials reveal past geological events, including intensive volcanic/impact degassing and resurfacing, as well as processes that occurred under wetter conditions.

10.10.2 Observational Constraints on Venus and Mars

This section presents data on surface materials and atmospheres in the context of gas–solid-type interactions. More information on planetary surfaces and atmospheres can be found in *Treatise on Geochemistry*, volume 1 (e.g., Fegley, 2003; McSween, 2003).

10.10.2.1 Physical and Chemical Signs of Surface Rock Alteration

Basalts are the more abundant igneous rocks on Mars and Venus. On Mars, the mafic composition of primary rocks is inferred from studies of Martian meteorites, *in situ* analyses by the Viking and Mars Exploration Rover (MER) missions, and remote sensing data (McSween, 2003; Rieder *et al.*, 2004;

McSween *et al.*, 2006). The mainly basaltic composition of Venus' plains is inferred from the morphology of widespread volcanic features observed in Venera 15–16 (Barsukov *et al.*, 1986) and Magellan (Crumpler *et al.*, 1997) radar images and X-ray fluorescence (XRF) analysis at the landing sites of Venera 13–14 and Vega 2 (Table 2). Gamma-ray spectrometric analysis of K, U, and Th at Venera 9–10 and Vega 1–2 landing sites are consistent with mafic rocks, while Venera 8 materials resemble alkaline rocks (Surkov *et al.*, 1987).

Despite their chiefly basaltic composition, surface rocks are physically and chemically altered. Planetary surfaces are partially covered by wind-blown material, rock fragments, and fine-grained soil. Porous layered deposits have been observed at the landing sites on Venus (e.g., Florensky *et al.*, 1977, 1983; Garvin *et al.*, 1984; Basilevsky *et al.*, 1985), at Meridiani Planum on Mars (Squyres *et al.*, 2006), and at the Columbia Hills in Gusev crater (Figures 1 and 2). It has been suggested that layered rocks at these sites were deposited from the atmosphere by eolian activity, explosive volcanism, and/or impact events (Garvin *et al.*, 1984; Basilevsky *et al.*, 1985, 2004; Grotzinger *et al.*, 2005; Knauth *et al.*, 2005).

Layered rocks observed at Venera 13–14 landing sites are mechanically weak materials with high porosities (50–60%), low densities ($1.4\text{--}1.5 \text{ g cm}^{-3}$, Florensky *et al.* (1983)) and low bearing capacities ($4\text{--}5 \text{ kg cm}^{-2}$, Avduevsky *et al.* (1983); $2.6\text{--}10 \text{ kg cm}^{-2}$, Kemurdzhian *et al.* (1983)). These rocks resist drilling similar to weathered porous basalt or ashy tuff (Barmin and Shevchenko, 1983). Layered rocks observed at the Venera 9–10 landing sites have

higher bearing capacities (30–300 kg cm^{-2}), although they possess properties analogous to volcanic tuffs and/or weathered basalts (Kemurdzhian *et al.*, 1983; Basilevsky *et al.*, 2004). Predominantly horizontal layering and some signs of cross-bedding (Venera 10) are consistent with deposition from the atmosphere followed by mild lithification and erosion. Fine-grained material observed in Venera 9–10 and 13 panoramas is the likely product of the degradation of local rocks (Florensky *et al.*, 1977; Garvin *et al.*, 1984; Basilevsky *et al.*, 2004).

Radar observations of Venus' volcanic plains by the Pioneer Venus, Venera 15–16, and Magellan orbiters revealed a dielectric constant (δ) of ~ 5 , typical for basalts (Ford and Pettengill, 1983; Pettengill *et al.*, 1997). These data are consistent with the electrical resistance of rocks at the landing site of Vega 2 ($10^6 \Omega \text{ m}$), which is common for heated basalts (Kemurdzhian *et al.*, 1983). However, rocks at the landing sites of Venera 13–14 had significantly lower electrical resistances (89 and $73 \Omega \text{ m}$, respectively; Kemurdzhian *et al.* (1983)). These measurements suggest lateral heterogeneity of surface material. On the highlands, $\sim 4.5 \text{ km}$ above the mean planetary radius (6051.5 km), surface materials are characterized by elevated radar reflectivity that corresponds to a dielectric constant as high as 20–30 (Ford and Pettengill, 1983; Pettengill *et al.*, 1997). These high δ values are inconsistent with unaltered igneous rocks.

Martian layered rocks at Meridiani Planum (Figure 2) and Gusev crater (e.g., Home Plate in the Columbia Hills) are also characterized by high porosity, low density, and sometimes cross-bedded

Table 2 Chemical composition of Venus surface materials at the Venera and Vega landing sites (mass %)

Oxide	Venera 13	Venera 14	Vega 2
SiO ₂	45.1 ± 3.0	48.7 ± 3.6	45.6 ± 3.2
Al ₂ O ₃	15.8 ± 3.0	17.9 ± 2.6	16.0 ± 1.8
FeO	9.3 ± 2.2	8.8 ± 1.8	7.7 ± 1.1
MnO	0.2 ± 0.1	0.16 ± 0.08	0.14 ± 0.12
MgO	11.4 ± 6.2	8.1 ± 3.3	11.5 ± 3.7
CaO	7.1 ± 0.96	10.3 ± 1.2	7.5 ± 0.7
K ₂ O	4.0 ± 0.63	0.2 ± 0.07	0.1 ± 0.08
TiO ₂	1.59 ± 0.45	1.25 ± 0.41	0.2 ± 0.1
SO ₃	1.62 ± 1.0	0.88 ± 0.77	4.7 ± 1.5
Cl	<0.3	<0.4	<0.3
Total	96.1	96.3	93.4

The data were obtained with XRF analysis. Uncertainties are $\pm 1\sigma$. All Fe is given as FeO. Note that Na content has not been measured. Sources: Surkov *et al.* (1984), Surkov *et al.* (1986).



Figure 1 Layered rocks and rock fragments at the surface of Venus at the landing site of Venera 13. The surface is black at Venus' conditions and reddish color represents the surface at room temperature (cf. Pieters et al., 1986). The upper part of the image has been artistically created by Don Mitchell using other Venus' surface images.



Figure 2 Layered sedimentary rocks at the Meridiani Planum, Mars. The rind-like features appear to cap the rocks and are probably more resistant to wind erosion than surrounding rocks. The image was taken with the Opportunity rover during 552 sol (13 Aug. 2005). Photo credit NASA/JPL/Cornell.

structures (e.g., Grotzinger et al., 2005; Fergason et al., 2006). Thermal inertia of soils and bedforms at both MER landing sites is consistent with particle diameters between 45 and 160 μm (silt to fine sand) (Fergason et al., 2006), which agrees with the grain size of the most easily suspended particles (\sim 100–150 μm , Greeley et al. (1980)). The finest grains could represent a globally homogenized material that has also been observed at the Viking and Mars Pathfinder landing sites (Christensen and Moore, 1992; Bell

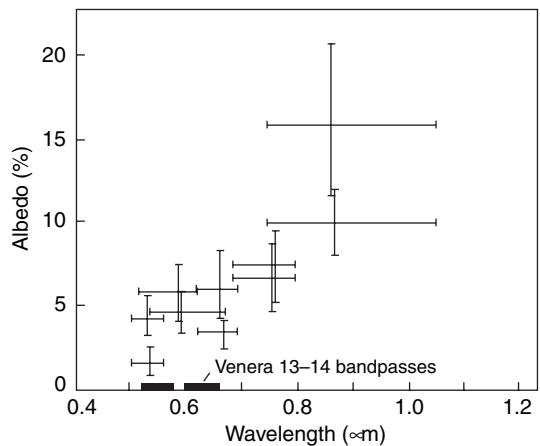


Figure 3 Reflectance properties of the Venus' surface at the Venera 9–10 landing sites (Ekonomov et al., 1980; Pieters et al., 1986). Horizontal bars indicate the width at half-height of the Venera filters. Reproduced from Pieters CM, Head JW, Patterson W, et al. (1986) The color of the surface of Venus. *Science* 234: 1379–1383. Reprinted from AAAs.

et al., 2000). In addition, rocks and soils are at least partially coated by particles with a diameter of 0.2–10 μm , which represents atmospheric dust.

On Mars and Venus, surface materials exhibit specific absorptions in the visible and near-infrared spectral ranges, indicating the presence of Fe^{3+} -bearing (ferric) species, which are uncommon in unaltered basalts. Reflectance spectra of Venus' surface measured at the Venera 9–10 landing sites (Ekonomov et al., 1980) roughly match with the spectrum of the ferric oxide hematite ($\alpha\text{-Fe}_2\text{O}_3$) heated to Venus' surface temperatures (Pieters et al., 1986), as can be seen in Figures 3 and 4. On Mars, the adsorption of light by ferric oxides is responsible for the reddish hue of dust, soils, and rock coatings (Soderblom, 1992; Bell et al. 2000). This inference is consistent with the detection of goethite ($\alpha\text{-FeOOH}$), hematite, nanophase Fe^{3+} oxides, and ferric sulfates and clays by thermal emission (Christensen et al., 2001; Glotch et al., 2006) and Mössbauer (Klingelhöfer et al., 2004; Morris et al., 2006) spectroscopies. Only minor alteration observed in Martian meteorites (Gooding, 1992; Bridges et al., 2001), which were excavated by impacts, implies that oxidation occurs mainly at the Martian surface.

Chemical analyses of surface materials on both planets reveal incorporation of volatile elements into alteration products. Venus' surface rocks (Table 2) are significantly enriched in S compared

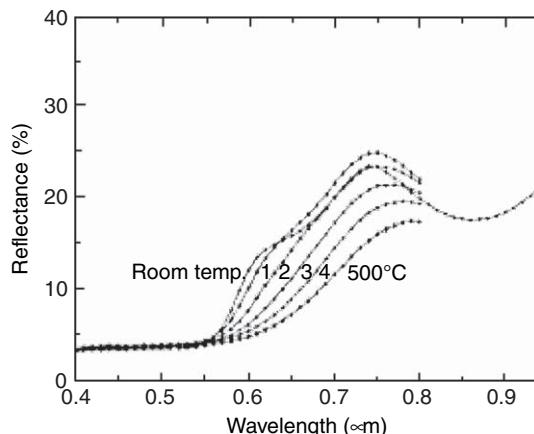


Figure 4 Laboratory reflectance spectra of the ferric oxide hematite at temperatures ranging from room temperature to 500°C. Reproduced from Pieters CM, Head JW, Patterson W, et al. (1986) The color of the surface of Venus. *Science* 234: 1379–1383. Reprinted from AAAs.

to basalts, which typically contain <0.2 wt.% S. Martian soils (**Table 3**) are enriched in S, Cl, Br, and bound H₂O compared to normal basalts. Rock fragments possess enrichments in S and halogens, indicating dust coatings and/or weathering rinds

(Rieder *et al.*, 1997, 2004; Gellert *et al.*, 2006). The variable abundances of volatiles imply their exogenic source. Elevated abundances of sulfur observed in Venus' and Mars' surface rocks may indicate global processes of basalt enrichment.

10.10.2.2 Atmospheric Compositions in the Context of Atmosphere–Surface Interactions

Carbon dioxide and N₂ are the major atmospheric components on Mars and Venus (**Table 1**). CO₂ is a major component of terrestrial volcanic gases (Symonds *et al.*, 1994) and could be an abundant product of mantle outgassing on other terrestrial planets. On Venus, the mass of atmospheric CO₂ is similar to the mass of CO₂ stored in terrestrial crustal carbonates (**Table 4**). This implies that the majority of CO₂ degassed on Venus resides in its atmosphere. On Mars, the low mass of CO₂ in the atmosphere, even when normalized to planetary mass, may indicate early incorporation of CO₂ in carbonates and/or escape of CO₂ into space.

In both atmospheres, active trace species play a major role in atmospheric chemical processes and rock alteration. On Mars, photochemically generated

Table 3 Chemical composition of martian soils detected *in situ* (in mass %)

Oxide	Viking 1, Chryse Planitia	Viking 2, Utopia Planitia	Average Viking soil	Average Mars Pathfinder	Opportunity rover, Meridiani Planum	Average Spirit soil, Gusev crater
SiO ₂	43	43	43.4 ⁺⁶ ₋₆	48.6 ± 2.5	45.5 ± 0.4	45.8 ± 0.44
Al ₂ O ₃	7.3	7 ^b	7.2 ⁺⁴ ₋₄	8.3 ± 0.8	8.8 ± 0.2	10.0 ± 0.22
Fe ₂ O ₃	18.5	17.8	18.2 ⁺⁵ ₋₂	17.5 ± 1.8	20.1 ± 0.2 ^a	15.8 ± 0.36 ^a
MnO					0.4 ± 0.02	0.31 ± 0.02
MgO	6	6 ^b	6.0 ⁺⁵ ₋₃	7.5 ± 1.2	7.2 ± 0.2	9.3 ± 0.23
CaO	5.9	5.7	5.8 ⁺² ₋₂	6.3 ± 1.0	7.52 ± 0.05	6.1 ± 0.27
Na ₂ O					2.2 ± 0.9	1.4 ± 0.3
K ₂ O	<0.15	<0.15	<0.15	0.3 ± 0.1	0.48 ± 0.1	0.41 ± 0.03
TiO ₂	0.66	0.56	0.6 ^{+0.25} _{-0.25}	1.1 ± 0.2	1.09 ± 0.05	0.81 ± 0.08
Cr ₂ O ₃					0.52 ± 0.02	0.35 ± 0.07
P ₂ O ₅					0.82 ± 0.04	0.84 ± 0.07
SO ₃	6.6	8.1	7.4 ⁺⁶ ₋₂	5.4 ± 1.1	4.93 ± 0.05	5.82 ± 0.86
Cl	0.7	0.5	0.8 ^{+1.5} _{-0.5}	0.6 ± 0.2	0.43 ± 0.03	0.53 ± 0.13
Br, ppm ^c					30 ± 30	40 ± 30
Ni, ppm					640 ± 40	450 ± 120
Zn, ppm					280 ± 40	300 ± 80
Total	88.7	88.7	89.4	97.8	99.2	99.4

^aFor Opportunity and Spirit data, all Fe is reported as FeO.

^bNot measured and assumed to be the same as in Viking 1 soils.

^cppm, parts per million by mass.

Sources: Clark *et al.* (1982), Bell *et al.* (2000), Rieder *et al.* (1997), Rieder *et al.* (2004), Gellert *et al.* (2004), Gellert *et al.* (2006). Viking data represent surface fines from Banin *et al.* (1992). Average Viking and Mars Pathfinder soils are from Bell *et al.* (2000). Opportunity data are for the first soil analysis (Rieder *et al.*, 2004). Spirit data are for average soil (Gellert *et al.*, 2004).

Table 4 Masses of volatiles in outer shells of Venus, Mars and Earth (g)

Gas	Venus atmosphere	Mars atmosphere	Earth atmosphere, hydrosphere and crust ^a
CO ₂	4.6×10^{23}	2.1×10^{19}	4.1×10^{23}
N ₂	1.1×10^{22}	7.6×10^{17}	5.3×10^{21}
H ₂ O	7.0×10^{18}	6.5×10^{15}	2.3×10^{24}
S	5.4×10^{19}	<2 × 10 ¹²	2.6×10^{22}
Cl	1.6×10^{17}	<2 × 10 ¹²	3.6×10^{22}
Abundances normalized to planetary masses (g/g)			
CO ₂	9.45×10^{-5}	3.3×10^{-8}	6.86×10^{-5}
N ₂	2.3×10^{-6}	1.2×10^{-9}	8.9×10^{-7}
H ₂ O	1.4×10^{-9}	1.0×10^{-11}	3.9×10^{-4}
S	1.1×10^{-8}	<3 × 10 ⁻¹⁵	4.4×10^{-6}
Cl	3.3×10^{-11}	<3 × 10 ⁻¹⁵	6.6×10^{-6}

^aTerrestrial data are from Ronov and Yaroshevsky (1976) and Volkov (1992).

species (e.g., CO, O₂, H₂, H₂O₂, O₃, HO₂, O, OH, H, NO, and NO₂) distributed throughout the lower atmosphere and can affect the surface (McElroy *et al.*, 1977; Yung and DeMore, 1999). On Venus, highly oxidized photochemically derived species are subjected to thermochemical reactions below clouds (48–60 km) and may not reach the surface. Near the surface, atmospheric chemistry is controlled by thermochemical reactions, with an exception of photochemical interaction between S₃ and S₄ gases. In contrast, in the Martian atmosphere, thermochemical processes are insignificant because of low densities and temperatures.

On Mars, densities of atmospheric gases vary with season and latitude. Nonsymmetric seasonal condensation of atmospheric CO₂ causes variations in total atmospheric pressure (James *et al.*, 1992). Similarly, the abundance and partial pressure (*P*) of water vapor is strongly dependent on season and latitude, reaching the highest values in spring in the Northern Hemisphere (Jakosky and Haberle, 1992).

On Venus, altitudinal variations in concentrations of trace gases are caused by temperature-pressure-dependent thermochemical reactions and Hadley circulation beneath the cloud deck. Thermochemical processes below the clouds create concentration gradients for at least several gases. For example, thermochemical interaction of CO with S-bearing gases to form COS could be implicated by the observed decreasing CO and increasing COS abundances toward the surface (e.g., Krasnopolsky and Pollack, 1994). The upward motion of gases in the equatorial region and downwellings in pole regions in the Hadley

cell could produce higher concentrations of CO in high latitudes and elevated abundances of COS in equatorial regions.

At Venus' surface, catalysis by solid phases and sluggish air circulation in the boundary layer may allow thermochemical equilibration among at least some gases. For example, a model of chemical equilibration among CO₂, CO, COS, and SO₂ at the surface is consistent with observations (Krasnopolsky and Pollack, 1994; Zolotov, 1996; Fegley *et al.*, 1997b). Concentrations of several other near-surface trace gases may be controlled by thermochemical equilibria as well (Table 5). The possibility of equilibration among gases increases below the surface, where migration of gases is controlled by diffusion through pore spaces. High porosity facilitates gas–surface interactions in the near-surface

Table 5 Composition of the near-surface atmosphere of Venus according to a thermochemical equilibrium model for conditions at the modal planetary radius (6051.4 km, 740 K, and 95.6 bar), in volume fraction

Gas	Mole ratio	Gas	Mole ratio
CO ₂	0.965	H ₂ S	1.1×10^{-7}
N ₂	0.035	H ₂	3.5×10^{-9}
SO ₂	1.3×10^{-4}	S ₃	8.5×10^{-10}
H ₂ O	3.0×10^{-5}	CS ₂	6.5×10^{-11}
COS	2.8×10^{-5}	SO	3.1×10^{-11}
CO	1.7×10^{-5}	S ₄	7.0×10^{-12}
S ₂	2.3×10^{-7}	O ₂	$10^{-21.3}$

Source: Zolotov MYu (1996) A model for the thermal equilibrium of the surface venusian atmosphere. *Geochemistry International* 33(10): 80–100.

layer. Without new data from Venus, assuming thermochemical equilibrium provides a way to evaluate concentrations of undetected gases.

Above Venus' surface, especially at elevations of several kilometers, equilibrium predictions are not consistent with observations (e.g., Krasnopolsky and Pollack, 1994; Zolotov, 1996). Disequilibria among major chemically active gases (CO_2 , SO_2 , CO , COS) in Venus' highlands constrains the evaluation of atmospheric composition and pathways of gas–mineral interactions.

10.10.3 Chemistry of Atmosphere–Surface Reactions on Mars and Venus

Atmospheric–surface reactions have been considered since the early 1960s in attempts to interpret space-craft and telescopic observations, as reviewed by Von Zahn *et al.* (1983), Lewis and Prinn (1984), Volkov *et al.* (1986), Banin *et al.* (1992), Fegley and Treiman (1992), Gooding *et al.* (1992), Zolotov and Volkov (1992), Fegley *et al.* (1997a), Wood (1997), and Fegley (2003). Despite many theoretical and some experimental efforts, there is no consensus among planetologists on the stability of minerals, as well as rates and timing of processes.

10.10.3.1 Redox Processes

Chemical reactions involving electron transfers (redox processes) are the major class of reactions at the atmosphere–surface interface on Venus and Mars. Photochemical production of oxidized atmospheric gases (e.g., SO_3 , O_2 , O_3) and volcanic supply of CO_2 , SO_2 , and H_2O create the potential for oxidation of freshly exposed igneous rocks. In fact, the observed oxidizing nature of surface materials on Mars and Venus (see Section 10.10.2.1) implies that atmospheric oxidants were present and involved. On Mars and Venus, surface rocks are only partially oxidized owing to limited masses of atmospheric oxidants compared to the amount of exposed rocks. On both planets, a major phase of mineral oxidation could have occurred during earlier epochs that were warmer and more humid (see Section 10.10.5).

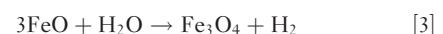
10.10.3.1.1 Oxidation on Venus

The atmospheric redox state near the mean planetary radius on Venus corresponds to conditions at which the iron oxides magnetite ($\text{Fe}^{2+}\text{Fe}_2^{3+}\text{O}_4$, Fe_3O_4)

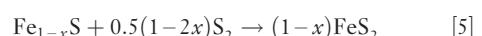
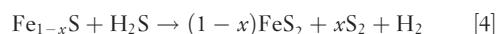
and hematite coexist with each other (Zolotov, 1996; Fegley *et al.*, 1997b), consistent with the interpretation of near-infrared reflectance patterns of rocks observed at the Venera 9–10 landing sites (Pieters *et al.*, 1986). An oxidized, hematite-bearing surface environment also agrees with results from the color experiments on the Venera 13–14 landers that provided a lower limit for atmospheric CO of 2.3–6.8 ppm (Zolotov, 1996; Fegley *et al.*, 1997b).

The strong absorption of light in the visible spectral range observed for Venera 9–10 and Venera 13–14 rocks (Figure 3) is consistent with the low albedo of magnetite and hematite at the surface temperature. However, the lack of Fe enrichment in analyzed rocks (Table 2) compared to typical basalts implies only a low-albedo coating of ferric oxides. An analysis of Venera 13 color data reveals an enrichment of surface fines in ‘red’ Fe^{3+} -bearing material (Shkuratov *et al.*, 1987). This interpretation agrees with diffusion-limited oxidation by atmospheric gases by which smaller grains alter faster. In addition, glassy impact-generated components in Venus’ fines may experience higher rates of alteration compared to crystalline minerals.

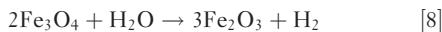
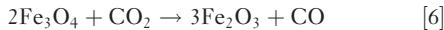
Although magnetite has not been detected on Venus, it can form as an intermediate phase during oxidation of ferrous silicates to hematite. Theoretical studies of mineral stability show that primary ferrous minerals in basalts (olivine, pyroxenes) could be oxidized by O-bearing atmospheric gases, leading to production of reduced gases. The pathways of oxidation to magnetite can be exemplified by the following reactions:



Here, ferrous iron oxide is a component of rock-forming silicates and silicate glasses. In addition to ferrous silicates, the primary iron sulfides troilite (FeS) and pyrrhotite (Fe_{1-x}S) can be oxidized to ferric oxides and/or pyrite (FeS_2). Pyrrhotite-to-pyrite conversions can occur through the uptake of H_2S and/or S_2 via reactions



The correspondence of atmospheric redox conditions to the conditions of the magnetite–hematite equilibrium is not likely to be accidental and could reflect ongoing magnetite-to-hematite transition via some of the following reactions:



It is not clear which atmospheric gas plays the major role in rock oxidation. If equilibrium models for atmospheric chemistry are valid, considered oxidation reactions have the same thermodynamic affinity, so the efficiency of each pathway is governed by kinetics. The efficiency of basalt oxidation by CO_2 at Venus' surface conditions has been demonstrated experimentally (Fegley *et al.*, 1995). Oxidation by SO_2 can occur through disproportionation of sulfur and release of S_2 (Terradellas and Bonnetain, 1973). Water is also a powerful oxidant of ferrous silicates in O_2 -free conditions. More experimental efforts are needed to evaluate the relative role of potential oxidation reactions. Note, at different altitudes, temperature and compositional changes should affect the thermodynamics and kinetics of redox reactions. New observations and experiments are needed to understand redox processes at high altitudes.

The coexistence of magnetite and hematite at the surface may buffer atmospheric CO_2/CO , SO_2/COS , SO_2/S_2 , and $\text{H}_2\text{O}/\text{H}_2$ ratios, which designate the oxidation state of the lower atmosphere. Buffering by reactions [6]–[8] means that the gas fugacity ratios remain unchanged until magnetite is consumed. Note that the magnetite-to-hematite conversion can occur through only one or two reactions followed by thermochemical re-equilibration among other gases. The dominance of ferrous minerals in mafic crustal rocks compared to the amount of oxidized gases and a slow rate of H_2 escape (Donahue *et al.*, 1997) could have prevented profound oxidation of rocks to magnetite and hematite. In other words, slowly reacting ferric oxides may maintain the redox conditions in the near-surface atmosphere, where the majority of air mass is located.

10.10.3.1.2 Oxidation on Mars

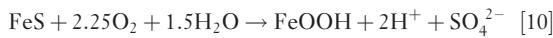
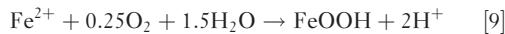
The reddish color of the Martian surface and the detection of hematite, goethite, nontronite (a ferric

smectite), and Ca-, Mg-, Fe^{3+} -sulfates by MER (Klingelhöfer *et al.*, 2004; Morris *et al.*, 2006; Glotch *et al.*, 2006) and Mars Express' OMEGA near-infrared spectrometer (Poulet *et al.*, 2005) is consistent with near-surface oxidation. As on Venus, Martian surface conditions thermodynamically favor oxidation of Fe^{2+} silicates and sulfides to Fe^{3+} -species and sulfates (O'Connor, 1968; Gooding, 1978; Gooding *et al.*, 1992). On Mars, oxidation of minerals may be driven by photochemically produced O_2 , O_3 , and H_2O_2 (e.g., Parkinson and Hunten, 1972; McElroy *et al.*, 1977). In addition, peroxides can form on surfaces of ultraviolet-irradiated mineral grains (Yen *et al.*, 2000), as well as by electrochemical processes in Martian dust devils and storms (Atreya *et al.*, 2004). In all cases, peroxides are produced via dissociation of atmospheric O-bearing gases. The presence of H_2O_2 and possibly other photochemically derived oxidants in the soil is consistent with results of Viking biological experiments (Oyama and Berdahl, 1977, 1979; Klein, 1978; Zent and McKay, 1994).

The mechanisms and environments of Martian surface oxidation could have varied in time and space and remain poorly known. Huguenin (1973a, 1973b) reported experimental results showing rapid photooxidation of magnetite to hematite in an O_2 -bearing atmosphere. However, subsequent experiments by Clark *et al.* (1979) and Morris and Lauer (1980) failed to observe alteration of magnetite, as well as photo-stimulated alteration of other minerals. This inconsistency questioned results from theoretical works in which the effects of solar irradiation on formation of goethite, clays, and other hydrated minerals were investigated (Huguenin, 1974, 1976). Similarly, the formation of peroxides in Martian soil through frost weathering of ferrous silicates (Huguenin *et al.*, 1979; Huguenin, 1982) remains to be verified. In contrast to inorganic species, photochemical oxidation of organic compounds occurs readily under Martian conditions, consistent with lack of detection of meteoritic and biological organic material in Martian surface materials.

The occurrence of hematite-rich spherules at Meridiani Planum (Figure 2; Klingelhöfer *et al.*, 2004), layered sulfate deposits at Valles Marineris and other locations (Gendrin *et al.*, 2005; Langevin *et al.*, 2005), and localized goethite and sulfate enrichments in Gusev crater (Ming *et al.*, 2006) reveal aqueous precipitation. It is possible that atmospheric and soil oxidants dissolved in surface aqueous

solutions and oxidized reduced solutes (e.g., Fe^{2+}) and sulfide minerals via overall reactions



where FeS represents iron sulfides and FeOOH stands for ferric oxides and oxyhydroxides. The presence of even traces of dissolved H_2O_2 may significantly increase the rate of oxidation. The contribution of H_2O_2 to aqueous oxidation might have been essential at low water/rock ratios that may represent episodic formation of water films or precipitation of atmospheric aerosols. However, the limited mass of atmospheric H_2O_2 ($\sim 6 \times 10^8 \text{ kg}$) and its rapid decomposition to O_2 and H_2O upon dissolution would limit the role of H_2O_2 in long-standing surface waters, which could have been typical during the earliest period of Martian history.

Reactions [9] and [10] produce protons that decrease solution pH and the oxidation rate. In terrestrial environments, aqueous oxidation of sulfides leads to pH values as low as -2 (Nordstrom *et al.*, 2000). On Mars, mineral precipitation in acidic near-surface aqueous environments is consistent with MER observations at Meridiani Planum and Gusev crater (Clark *et al.*, 2005; Haskin *et al.*, 2005; Ming *et al.*, 2006). In particular, the detection of jarosite (a complex Fe^{3+} sulfate), in layered rocks at Meridiani, indicates low-pH, oxidizing conditions at the time of mineral formation (Klingelhöfer *et al.*, 2004). The presence of jarosite also shows that limited amounts of solution did not neutralize before evaporating and/or freezing. In other locations, surface aqueous processes could have occurred at higher water/rock ratios and led to complete neutralization, consistent with the detection of smectites in Noachian deposits (Poulet *et al.*, 2005). Note that the lack of atmospheric oxidants could have caused subsurface solutions to be reduced and alkaline in composition.

The scale of aqueous oxidation on Mars could have been limited by the mass and production rate of atmospheric oxidants. Periods of intensive volcanism may have been characterized by low concentrations of oxidants owing to their interaction with reduced volcanic gases (Catling and Moore, 2003). Although strong impacts could have generated an array of strong oxidants (see Section 10.10.3.8), the limited oxidation of surface materials is consistent with the widespread occurrence of olivine and

Fe^{2+} -bearing pyroxenes in soils and igneous rocks (Hoefen *et al.*, 2003; Klingelhöfer *et al.*, 2004; Mustard *et al.*, 2005). Spectral and chemical observations of rock fragments and red-colored magnetite-rich dust collected on MER magnets (Goetz *et al.*, 2005) imply that formation of ferric oxides affected only the surfaces of mineral grains and rock fragments.

During periods characterized by a reducing but warm atmosphere, surface oxidation could have occurred by water via reactions [3] and [8]. Escape of H_2 produced through dissociation of H_2O would have contributed to net oxidation. Deep below the Martian surface, oxidation by liquid water may have occurred over Mars' history. Slow interaction of H_2 with dissolved carbonate species (HCO_3^- , CO_3^{2-}) can account for traces of CH_4 in the current atmosphere (Table 1).

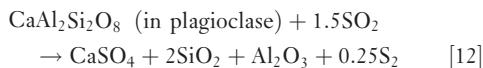
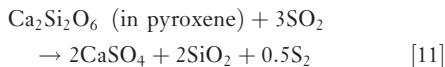
10.10.3.2 Reactions with Sulfur Gases

Sulfur is abundant in the Martian soil, rock coatings, and sedimentary deposits, as well as in Venus rocks in which S was analyzed (Tables 2 and 3). In all cases, S content does not correlate with rock composition, consistent with a secondary origin of sulfur. Despite elevated S content, only a fraction of metals (Ca, Mg, Fe) is bonded with sulfur, indicating incomplete alteration with respect to sulfur.

Although the mineralogy of S-bearing phases remains to be determined for many of Mars' and all of Venus' rocks, sulfates are the most likely secondary minerals. On Mars, the presence of sulfates is inferred from correlations with Mg (e.g., Clark *et al.*, 1982, 2005; Bell *et al.*, 2000), from infrared observations by MER (Glotch *et al.*, 2006), and from near-infrared spectra from the Mars Express orbiter (Gendrin *et al.*, 2005; Langevin *et al.*, 2005). These data are consistent with the presence of Mg sulfates with various degrees of hydration (e.g., kieserite, $\text{MgSO}_4 \cdot \text{H}_2\text{O}$, epsomite, $\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$), gypsum ($\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$), and ferric sulfates. Gypsum is also present in some Martian meteorites (Bridges *et al.*, 2001). Several other sulfates may be present as well.

On Venus, thermodynamic instability of rock-forming calcium silicates (plagioclase, pyroxenes) with respect to atmospheric SO_2 (Table 1) should drive the formation of anhydrite (CaSO_4) (Barsukov *et al.*, 1982; Volkov *et al.*, 1986; Fegley and Treiman, 1992). Formation of sulfates in O_2 -free conditions can

occur through disproportionation of SO_2 via gas–solid-type reactions,



Alternatively or additionally, oxidation of S^{4+} to S^{6+} by atmospheric CO_2 (e.g., Fegley and Treiman, 1992) or H_2O may form CaSO_4 , releasing CO or H_2 , respectively. Although the presence of anhydrite in Venus' rocks has not been confirmed, conversion of diopside ($\text{CaMgSi}_2\text{O}_6$) to CaSO_4 has been observed in modeling experiments at a temperature of 1106 K (Fegley and Prinn, 1989).

Despite a strong thermodynamic drive for anhydrite formation, the composition of Venus' rocks (**Table 2**) indicates incomplete sulfurization of Ca silicates. This is illustrated in **Figure 5**, showing sulfur content in rocks compared to anhydrite. It can be seen that Venera 14 rocks are less altered with respect to S consumption. This is consistent with reduced physical weathering relative to Venera 13 rocks (Garvin *et al.*, 1984; Basilevsky *et al.*, 1985), which suggests younger rocks at the Venera 14 landing site.

In addition to sulfates, surface materials on both planets may contain secondary sulfides formed at the surface. On Mars, cooling of H_2S -rich hydrothermal fluids may cause precipitation of pyrite, as happens on Earth in oceanic ‘black smokers’. Rare occurrences of liquid water at the surface and limited oxidants could have restricted subsequent oxidation of pyrite to ferric oxides and sulfates. In other words, both primary (pyrrhotite) and secondary (pyrite) sulfides

may be present in Martian surface materials. Likewise, thermodynamically unstable olivine and magnetite are present in Martian soils and dust (Hoefen *et al.*, 2003; Klingelhöfer *et al.*, 2004; Goetz *et al.*, 2005; Morris *et al.*, 2006).

On Venus, instability of pyrrhotite and troilite with respect to oxidation (see Section 10.10.3.1) does not exclude the occurrence of pyrite, which can stably coexist with magnetite and hematite at the conditions of plains. Note, however, that evaluations of mineral stability are limited by the uncertainties of thermodynamic data for gas and solid phases, impure solid phases, and uncertainties in the composition of the near-surface atmosphere. These factors led to controversial conclusions about pyrite stability on Venus (e.g., Barsukov *et al.*, 1982; Klose *et al.* 1992; Fegley and Treiman, 1992; Wood, 1997; Fegley *et al.*, 1997a).

Excess sulfur in surface materials on Venus and Mars likely has a volcanic source. On Mars, sulfur outgassing is consistent with the isotopic composition of S in Martian meteorites, which reveals photochemistry-driven mass-independent fractionation of sulfur in the atmosphere (Farquhar *et al.*, 2000). On Venus, a volcanic origin of sulfur agrees with widespread volcanic features and elevated abundances of sulfur gases in the atmosphere. On both planets, volcanic and impact processes could have recycled sulfur through atmosphere and crustal reservoirs.

On Venus and Mars, photochemical oxidation of sulfur gases produces sulfuric acid aerosols. In the sulfur-rich Venusian atmosphere, constantly produced aerosols form clouds at an altitude of 48–60 km that do not affect the surface (Von Zahn *et al.*, 1983; Esposito *et al.*, 1997). On Mars, sulfuric acid aerosols form after volcanic eruptions (Settle, 1979) and impact cratering (see Section 10.10.3.8). The negligible amounts of S-bearing gases in the present atmosphere (**Table 1**) suggests that they were consumed by weathering processes that followed the last volcanic or impact event. Deposition of sulfate aerosols leads to their neutralization through mineral alteration. As a result, sulfates form on the surfaces of mineral grains and rock fragments. In addition to sulfates, acid weathering by volcanic/impact aerosols would cause deposition of silica and ferric oxides. Indeed, Fe^{3+} -rich coatings, likely nanophase oxides and oxyhydroxides, are ubiquitous on Martian surface materials (e.g., Morris *et al.* 2000, 2006; Goetz *et al.*, 2005) and silica coatings have been suggested from orbital thermal infrared spectra (Michalski *et al.*, 2005).

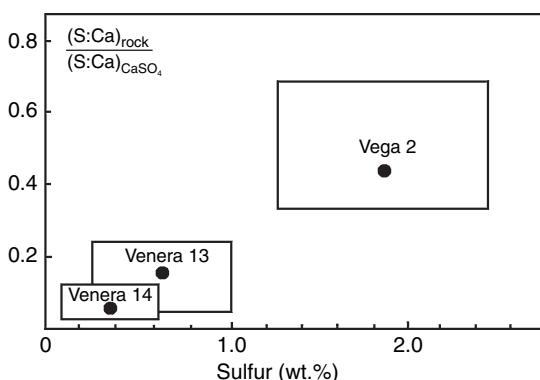


Figure 5 Sulfur content in Venus' rocks compared with that in Ca sulfate (anhydrite). Filled circles represent mean content and boxes show uncertainties. Data are from Surkov *et al.* (1984, 1986).

Although thermodynamics favors mineral weathering by sulfur gases under cold and dry Martian conditions (Gooding, 1978; Gooding *et al.*, 1992), most proposed reactions remain poorly explored. Modeling experiments (Clark *et al.*, 1979) indicate that conversions of carbonates and nitrates (if they are present) to corresponding sulfates are the most likely pathways of interaction of gaseous sulfur oxides with surface minerals. However, conversion of silicates to sulfates could be inhibited without aqueous phase.

10.10.3.3 Reactions with H₂O

Water vapor exists in the atmospheres of Mars and Venus (**Table 1**) and can interact with surface materials. On Mars, water condenses as ice (see Chapter 10.11), adsorbs on surface materials, and chemically bonds to minerals. On Venus, water vapor cannot condense anywhere near the surface and could be consumed via surface oxidation (e.g., reactions [3] and [8]) rather than through formation of OH-bearing minerals. On both planets, early atmospheres could have contained a higher abundance of water (Owen, 1992; Donahue *et al.*, 1997), causing competitive hydration and oxidation of surface materials. Low-temperature conditions at elevation would have favored hydration and restricted dehydration. Many minerals may have lost water during subsequent evolution.

On Mars, the occurrence of water ice and hydrated minerals is controlled by their thermodynamic stability and kinetics of hydration and dehydration. Stabilities of ice and hydrated minerals on the surface vary with latitude and season. Ice and highly hydrated minerals (e.g., epsomite, gypsum, antarcticite, CaCl₂·6H₂O) are stable at high latitudes and during winter in mid-latitudes. Elevated abundances of ground hydrogen at latitudes greater than ~40° inferred with neutron spectrometry from orbit (Boynton *et al.*, 2002; Feldman *et al.*, 2004; Mitrofanov *et al.*, 2004) agree with thermodynamic models of ice stability (e.g., Mellon and Jakosky, 1995). Hydrated minerals may also contribute to the high abundance of subsurface hydrogen (in ice + hydrates) at high latitudes. However, high H content in some large equatorial regions (e.g., Arabia) is inconsistent with thermodynamic models of the latitudinal distribution of ice and hydrates. It is possible that highly hydrated minerals formed at low latitudes during periods of high obliquity (Kieffer and Zent, 1992) and have not dehydrated to minerals stable under current climatic conditions (Feldman *et al.*, 2004). Slow dehydration of

epsomite, MgSO₄·7H₂O, provides an explanation. However, this interpretation has to be reconciled with signs of seasonal hydration and dehydration at mid-latitudes inferred from bound water absorption at 6.1μm (Kuzmin *et al.*, 2004).

At Venus' surface, all hydrated minerals and the majority of OH-bearing minerals are unstable (Zolotov *et al.*, 1997). Only some micas (e.g., phlogopite, KMg₃Al₃Si₂O₁₀(OH, F)₂) may be stable, especially at elevation. Partial substitution of the OH group by F favors phlogopite stability. Despite its stability, OH–F phlogopite formation requires transport of metals from different minerals, which is unlikely at present surface conditions. If micas and amphiboles were supplied from interior, they could survive at the current surface owing to slow dehydration rates (Johnson and Fegley, 2003, 2005). However, dehydration may have occurred at elevated atmospheric greenhouse temperatures caused by global volcanism ~0.5 Ga.

On both planets, the loss of water from the atmosphere in oxidation processes and H₂ escape could be compensated by internal processes. On Mars, polar caps, ground ice, and hydrated minerals are the major near-surface water reservoirs. Magmatic H₂O degassing could have played a role in earlier epochs. On Venus, atmospheric water could be resupplied via diffusion of H₂O and H₂ gases from nominally anhydrous minerals, which may account for the majority H₂O inventory (Zolotov *et al.*, 1997). However, volcanic degassing on Venus may not be an important source of atmospheric water because of suppressed explosive activity (Head and Wilson, 1986), a low abundance of H₂O in magmas, and its high solubility.

10.10.3.4 Reactions with CO₂

Carbon dioxide is the most abundant gas in the atmospheres of Mars and Venus that significantly affects mineral–gas interactions. On Venus, the massive CO₂ atmosphere accounts for the high surface pressure and temperature (see **Table 1**), and could be the major gas oxidizing surface materials (reactions [1] and [6]). On Mars, seasonal condensation and sublimation of CO₂ is responsible for the annual cycle of atmospheric pressure (James *et al.*, 1992; see also Chapter 10.11). The presence of carbonates in Martian meteorites reveals the fixation of atmospheric CO₂ in the crust.

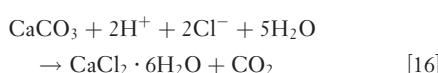
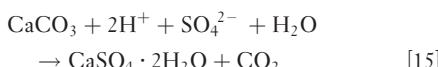
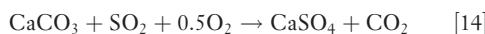
At present conditions, a significant chemical fixation of CO₂ in carbonates is unlikely on both planets. Experimental data demonstrate that carbonatization

of silicates is favored under aqueous conditions (Tanner *et al.*, 1985), which are absent at current surfaces. Although formation of Martian carbonates on irradiated surfaces has been predicted in theoretical models (Huguenin, 1974, 1976), an experimental work (Mukhin *et al.*, 1996) reveals the possibility of photo-stimulated decomposition of carbonates. Booth and Kieffer (1978) experimentally demonstrated formation of only traces of carbonates during interaction of Ca silicates with CO₂ under Martian temperatures with or without mild irradiation.

Although low totals in Venus surface analysis (**Table 2**) may reveal a presence of carbonates (Fegley *et al.*, 1997a), formation of calcite (CaCO₃) and dolomite (CaMg(CO₃)₂) is forbidden because of their thermodynamic instability with respect to high fugacity of atmospheric SO₂ (e.g., Volkov *et al.*, 1986; Fegley and Treiman, 1992), which would destroy them rapidly (Fegley and Prinn, 1989). If indigenously formed calcite is present on Venus' surface, the conversion of calcite to anhydrite may occur via net reaction



which was studied by Terradellas and Bonnetain (1973), and would result in rapid destruction of calcite at Venus's surface temperature. Siderite (FeCO₃) is unstable with respect to oxidation to ferric oxides. On the Martian surface, Ca carbonates could be attacked by SO₂, SO₃ (Clark *et al.*, 1979; Gooding *et al.*, 1992), and acid aerosols and rains of volcanic/impact origin, as illustrated by the reactions



Magnesite (MgCO₃) is the only rock-forming carbonate stable with respect to SO₂ on the surfaces of Venus and Mars. However, magnesite has not been firmly detected in the Martian soil and its occurrence at Venus' surface is unlikely because of thermodynamic and kinetic problems for its formation at present conditions.

The rapid destruction of calcite and dolomite compared to slow, if any, carbonatization of surface materials implies a likely absence of carbonates at the surface of Venus. An exposure of ancient sedimentary carbonates is not consistent with basaltic lava morphology of Venus's plains, and long sinuous channels

(Baker *et al.*, 1997) could have been formed by molten chlorides (see Section 10.10.3.5) rather than low-viscosity carbonate magmas. The paucity of carbonates on Venus agrees with the fact that there are similar masses of CO₂ in Venus' atmosphere and Earth's carbonates (**Table 4**). It follows that a popular idea of buffering atmospheric CO₂ by the Urey (1952) carbonate–silicate equilibrium,



(e.g., Adamchik and Draper, 1963; Mueller, 1964; Lewis, 1970; Fegley and Treiman, 1992), may be invalid despite the fact that the equilibrium P_{CO₂} in reaction [17] at ~740 K is similar to the P_{CO₂} at Venus' surface. Note, however, that a possibility for equilibrium [17] may exist below the surface (see Section 10.10.5).

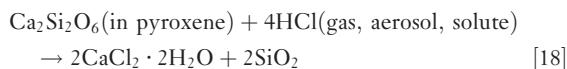
On Mars, secondary carbonates in crustal igneous rocks may account for the major reservoir of CO₂ formed in aqueous conditions. It is also possible that carbonates were deposited in a primordial ocean on Venus, if it ever existed. Subsequent volcanic and impact degassing of chemically active gases would have led to chemical weathering of surface carbonates on both planets. Sulfates and chlorides are the likely minerals formed after carbonates (reactions [13]–[16]). On Venus, early formed carbonates (if they formed) could be partially/completely decomposed, contributing to greenhouse heating from a thickening atmosphere (see also Section 10.10.5).

10.10.3.5 Reactions with Halogens

Halogens are delivered into planetary atmospheres through magmatic and impact degassing. On Earth (Symonds *et al.*, 1994) and Mars, they degas as hydrogen halides (HCl, HF) and SiF₄. On Venus, hydrogen-depleted magmas can emit diatomic halides, S- and S–O-halides, and SiF₄. The low abundances of Cl-, F-bearing gases in the atmospheres of Mars and Venus (**Table 1**) imply that degassed halogens have been consumed by surface materials. It is likely that planetary crusts are enriched in previously degassed halogens that are mainly stored in metal halides (e.g., NaCl, KCl, CaCl₂, MgCl₂) and halogen-bearing salts (apatite). Chlorides could be the most abundant halogen compounds owing to the higher solar abundance of Cl compared to other halogens. In addition, significant quantities of fluorine could be preserved in igneous rocks, especially those with elevated alkali content (Venera 8, Venera 13

landing sites, Columbia Hills basalts). On both planets, heating of surfaces and crusts could have caused mobilization, redistribution, and recycling of halogens. These processes may include magmatic/impact degassing, assimilation in magma chambers, melting and release of halogen-bearing salts, as well as aqueous activity and brine formation.

Martian surface materials are characterized by elevated and variable Cl and Br contents compared to basaltic rocks. Incorporation of degassed halogens in the soil could have occurred through precipitation of volcanic/impact aerosols, dissolution in surface aqueous solutions, and gas–mineral-type reactions. Hydrolysis of hydrogen halides produces acidic solutions that dissolve surface minerals followed by precipitation of halides of Ca, Mg, Na, K, and Fe upon freezing or evaporation. Weathering of rocks by acid gases, aerosols, acid rains, and surface solutions would convert primary and secondary minerals to corresponding halides. These transformations can be illustrated by reactions [16] and [18]:



Throughout Mars' history, these highly water-soluble salts were probably mobilized and redeposited by sporadic aqueous solutions. At present, deep crustal brines are most likely enriched in Cl and Br, which lower the freezing temperature (Burt and Knauth, 2002). Variable S/Cl/Br ratios in layered deposits at Meridiani Planum are consistent with element separation from aqueous processes (Clark *et al.*, 2005).

Hydrogen halides (HCl and HF) are present in trace amounts in Venus' atmosphere (Table 1). These gases could have originated from hydrolysis of H-free volcanic gases (e.g., Cl₂, SCl₂, SiF₄) and halogen-bearing surface materials (apatite, glasses). In turn, their interactions with primary and secondary minerals would lead to the formation of halogen salts. Although halogens have not been analyzed in Venus' rocks, the high upper limits of Cl (Table 2) do not exclude elevated abundances of halogens.

It is possible that atmospheric abundances of HCl and HF reflect chemical equilibria among these gases and secondary minerals at the surface (Mueller, 1968; Lewis, 1970; Fegley and Treiman, 1992). The low masses of atmospheric halogens (Table 4) compared to possible masses of halogens in a permeable surface layer imply that the atmospheric abundances of

halogens are controlled by surface mineralogy. In fact, atmospheric masses of Cl and F can be consumed in global layers 22 and 0.15 cm thick, respectively (at 0.1 wt.% of Cl and F). A comparison of the mass of Cl in Venus' atmosphere with that in terrestrial oceans and sediments (Table 4) suggests a comparable reservoir of halogens in Venus' crust (Fegley and Treiman, 1992; Zolotov and Volkov, 1992).

Venus' halides could have been mobilized in the crust because of their low melting points compared to silicates. Channels (Baker *et al.*, 1997) and crater outflows observed in Magellan radar images may have been formed by molten halides. It is possible that eutectic melts of halogen salts currently exist in Venus' crust. Likewise, Cl–Br–Ca–Mg–Na–(SO₄) eutectic brines may be present deep below the Martian surface.

10.10.3.6 Condensation of Metal and Nonmetal Species, the Role of Altitude

Volcanism, impacts, gas–solid-type weathering, and thermal rock outgassing delivered many elements into planetary atmospheres before they condensed. Terrestrial silicate volcanism degasses many metals (e.g., Na, K, Bi, Hg, Pb, Zn, Sb) and nonmetals (B, Se, As) (Symonds *et al.*, 1992). On Venus and Mars, volcanic sources of S in surface materials may also imply elevated abundances of Se, Te, As, and volatile metals.

A sizeable temperature gradient between Venus' plains (~730–750 K) and mountaintops (~650–700 K) should cause preferential condensation of heavy elements in elevated regions. Thermodynamic calculations demonstrate that sulfides, sulfosalts, halides, and chalcogenides of Pb, Bi, As, Cu, and Sb may condense on mountaintops and account for the high dielectric constant (Brackett *et al.*, 1995; Schaefer and Fegley, 2004). Lower temperatures at elevation also lead to lower rates of chemical weathering and would favor longer preservation of locally unstable phases. It follows that some metal/sulfide condensates may be unstable with respect to gas–solid weathering reactions.

10.10.3.7 Kinetic Aspects of Surface–Gas-Type Reactions

Gas–solid-type reactions affect individual mineral grains that are subjected to isochemical weathering. As a result, alteration products should consist of

mechanical mixtures of minerals formed through weathering of primary phases. Because weathering products form at the surface of mineral grains, smaller particles are subjected to more alteration. In many cases, weathering is controlled by diffusion of reacting gases through a layer of alteration products. Physical removal of alteration products from mineral surfaces would increase the rate of chemical weathering. It follows that small particles could be most chemically altered, consistent with observations on Mars (e.g., Yen *et al.*, 2005) and Venus (Shkuratov *et al.*, 1987). Despite wind abrasion and thermodynamic potential for weathering reactions at planetary surfaces, many gas–mineral-type interactions are inefficient because of slow reaction rates, especially at the diffusion-controlled stage. As a result, significant portions of surface rocks on Mars and Venus remain unaltered.

On Venus, kinetic parameters of key oxidation reactions ([1]–[3], [6]–[8]) and sulfurization reactions ([11] and [12]) remain unknown. Although rapid formation of anhydrite through SO_2 –calcite interaction is evident from many experiments (e.g., Terradellas and Bonnetain, 1973; Fegley and Prinn, 1989), this reaction may not be relevant to present-day Venus (see Section 10.10.3.4).

On Mars, rates and mechanisms of gas–solid weathering processes remain poorly explored, partially because of low reaction rates at subzero temperatures and low densities of atmospheric gases. Formation and reactivity of carbonates through gas–solid reactions (Booth and Kieffer, 1978; Mukhin *et al.*, 1996) remains to be confirmed, including effects of irradiation and gas diffusion through reaction products. Study of photo-stimulated oxidation, hydration (e.g., Huguenin, 1973a, 1973b, 1974, 1976, 1982; Morris and Lauer, 1980), and dehydration (Morris and Lauer, 1981; Yen *et al.*, 1999) could be revisited. Despite modeling experiments by Clark *et al.* (1979), kinetic parameters of low-temperature weathering by sulfur and hydrogen halide gases remain unknown. The mechanism of seasonal hydration and dehydration remains unknown as well.

10.10.3.8 Local Reactions Caused by Impacts and Volcanic Eruptions

Impacts and volcanic eruptions create highly variable local environments in which chemical transformations are governed by mass transfer and reaction rates. In impact events, thermal dissociation of species leads to chemically contrasting gas mixtures

(e.g., H_2 , O_2 , CO , SO_3 , nitrogen oxides) that affect local and global atmospheric environments (e.g., Toon *et al.*, 1997; Prinn and Fegley, 1987; Pope *et al.*, 1997). Reactions in volcanic/impact sites can include oxidation of ferrous iron and sulfides, reduction of iron to $\text{Fe}(0)$ metal at $T > 1500$ K, dehydration and hydration of minerals, thermal decomposition of carbonates, and conversion of carbonates to sulfates. On Earth, oxidation by volcanic H_2O often leads to hematite-bearing pumice with a reddish color. On Mars and Venus, high-temperature oxidation by H_2O and/or CO_2 in impact/volcanic gases could have contributed to the formation of sulfates and ferric oxides in soil and dust. Although hot CO_2 could be the major oxidizing agent on Venus (Fegley *et al.*, 1995), impacts of comets and carbonaceous chondrites may cause oxidation by H_2O . On both planets, hematite could have formed via oxidation in impact/volcanic events.

On Mars, slow cooling of lavas and impact melts from inefficient convection and low atmospheric density allowed a great degree of chemical alteration of melts and hot rocks. However, the low P_{O_2} in the atmosphere, especially during periods of volcanic activity, would not have favored rapid oxidation of hot surface materials. On Venus, rapid cooling of silicate melts at the contact with the dense atmosphere (Frenkel and Zabalueva, 1983) could have been compensated by the high densities of chemically active gases (CO_2 , SO_2).

On Mars, volcanic/impact melting of ice-rich deposits could have led to local hydration, especially in fluidized outflow deposits. Materials in Martian surge deposits could have been modified (oxidized, hydrated, weathered) shortly after deposition (Knauth *et al.*, 2005). On Venus, an impact-generated layered material suggested at Venera landing sites (e.g., Basilevsky *et al.*, 2004) may be composed of materials that were altered by impact events. In fact, the presence of chemically distinct fine-grained material in parabolic crater ejecta deposits is consistent with their elevated dielectric constant (up to 8) (Campbell *et al.*, 1992; Pettengill *et al.*, 1997).

On Earth, Mars, and Venus impact/volcanic degassing of sulfur-rich rocks would lead to an array of sulfur species and could affect global atmospheric composition. Impact decomposition of crustal sulfates during the Chicxulub event at the Cretaceous–Tertiary boundary released sulfur oxides into the atmosphere followed by downpours of acid rain (Pope *et al.*, 1997; Gupta *et al.*, 2001). On sulfate-rich Mars, raining from impact-generated

clouds (Segura *et al.*, 2002) could be responsible for observed signs of acid sulfate weathering (Zolotov and Mironenko, 2007). Thermal decomposition of carbonates during impacts and/or their interaction with released sulfur oxides and acid aerosols/rainfalls would have obliterated surface carbonates on Mars. On Venus, impact/volcanic release of sulfur gasses would temporarily affect the mass of sulfuric acid clouds.

10.10.4 Wind-Related Processes

Eolian processes provide the major mechanism for mass transfer on contemporary surfaces of Mars and Venus (Greeley and Iversen, 1985). Both erosional (yardangs) and depositional (dunes) features are observed (Greeley *et al.*, 1992, 1997). On Mars, dust devils and global dust storms are widely observed from orbit and the surface (Balme and Greeley, 2006). On both planets, the orientation of eolian features is consistent with general circulation models.

The effectiveness of eolian transport depends on air density, temperature, and gravity, which are significantly different on Mars and Venus (Greeley *et al.*, 1992, 1997). Nevertheless, models and experiments demonstrate that particles that are a few hundred microns in diameter can be transported on both planets. Particles with a diameter of $\sim 100 \mu\text{m}$ represent the most mobile fractions (Greeley *et al.*, 1980, 1992, 1997). Larger and smaller particles require stronger winds to be transported.

On Mars, low gravity and high wind velocity overcome the effects of low air density and cause significant eolian transport. Dust with grain sizes of a few microns is homogenized in the atmosphere on a planetary scale. Globally, deflation characterizes some low-albedo regions and accumulation is typical in high-albedo regions. Larger particles (up to perhaps a centimeter in diameter) are transported by saltation and creep, forming areas dominated by deflation and accumulation.

On Venus, measurements of wind speed near the surface ($<2 \text{ m s}^{-1}$, Ksanfomality *et al.* (1993)), extrapolation of wind velocities to the surface ($0.5\text{--}1 \text{ m s}^{-1}$, Counselman *et al.* (1979)), and wind tunnel experiments (Greeley *et al.*, 1987) suggest the possibility of present-day eolian activity of Venus. This is consistent with the observation of wind-related features (mostly wind streaks) in Magellan radar images (Greeley *et al.*, 1997). Fine-grained material is mainly supplied from impact cratering and chemical weathering. However, the apparent paucity of dunes on

Venus could be related to a lack of sand-size particles.

The orientation of Venusian wind features indicates the upper altitude super-rotation of the atmosphere and the extension of Hadley circulation to the surface (Greeley *et al.*, 1997). Wind transport is consistent with the presence of parabola-shaped zones of low radar backscatter around the relatively fresh impact craters. Models show that wind activity would disperse the inferred fine-grained materials in the parabolas over a few tens of millions years.

Without aqueous solutions, physical weathering of surface rocks is restricted to wind abrasion and rock disintegration caused by chemical weathering of individual grains. Despite vivid signs of eolian activity, the estimated rates of eolian erosion are small on both planets. On Mars, an erosion rate of $\sim 0.03 \text{ nm yr}^{-1}$ indicates slight surface changes since the Hesperian epoch ($\sim 3 \text{ Ga}$) (Golombek *et al.*, 2006). Modeling experiments demonstrate the possibility of rock abrasion even for moderate wind velocities ($0.5\text{--}0.7 \text{ m s}^{-1}$) at the surface of Venus (Greeley *et al.*, 1987). Despite these results, volcanic features remain well preserved (Crumpler *et al.*, 1997). In particular, during the Magellan mission, no surface changes were observed that could be attributed to eolian processes.

10.10.5 Atmosphere–Surface Interactions throughout History

Although gas–solid-type interactions prevailed over the majority of planetary histories, the effectiveness of these interactions varied over time along with changes in surface temperatures and atmospheric compositions. On Mars and Venus, significant atmosphere–surface interaction could have occurred earlier in history under wetter conditions and different temperatures, as well as through multiple episodes of volcanic eruptions and impacts.

Elevated D/H ratios in the current atmospheres of Mars ($7.8 \pm 0.3 \times 10^{-4}$) and Venus ($2.4 \pm 0.5 \times 10^{-2}$) compared to the Earth’s oceanic value (1.56×10^{-4}) indicate more humid conditions in their earlier atmospheres, which deteriorated by the escape of H (Owen, 1992; Donahue *et al.*, 1997; Krasnopolksky and Feldman, 2001). On Mars, a low $^{14}\text{N}/^{15}\text{N}$ ratio in the atmosphere (170 ± 15 , Owen (1992)) compared to the Earth’s value (272) implies gas escape from a denser early atmosphere. Despite a low solar luminosity, the greenhouse effect in denser and/or more humid atmospheres could have produced elevated

surface temperatures on Mars and Venus. Note, however, that cloud condensation should have affected gas abundances (e.g., CO₂ condensation on Mars, Kasting (1991)) and surface temperatures through albedo changes (Hashimoto and Abe, 2001).

Oxidation of ferrous minerals and sulfides by water (reactions [3] and [8]) and H₂ production is a likely process on early planets (Lewis and Prinn, 1984; Wänke and Dreibus, 1987). H₂ escape contributes to net oxidation of surfaces and atmospheres. The high oxidation state of most Martian meteorites is consistent with early oxidation of the lithosphere by H₂O.

Hydration of surface material is another likely process on early planets. On Mars, somewhat warmer and wetter conditions during the Noachian epoch are supported by the presence of phyllosilicates (Poulet *et al.*, 2005), valley networks (Carr, 1996), and heavily eroded terrains (Craddock and Howard, 2002). On Venus, the evolution of the greenhouse effect is poorly constrained and early hydration remains conjectured. The majority of hydrated phases would have been dehydrated during subsequent evolution, especially during increases in surface temperatures.

Humid conditions in the early atmospheres would have favored the fixation of atmospheric CO₂ in carbonates. On Mars, formation of carbonates in aqueous environments provides a major pathway to diminish an early CO₂ atmosphere (Pollack *et al.*, 1987). On Venus, carbonates could have been deposited in a primordial ocean, if one ever existed. Subsequent metamorphic reactions in the lithosphere could have led to the establishment of carbonate–silicate equilibria below the surface. Such a mechanism might be consistent with a buffering atmospheric CO₂ by equilibrium [17]. Note that Ca minerals below the surface should be protected from atmospheric SO₂, HCl, and HF.

Throughout the geologic histories of Mars and Venus, volcanic activity and impact events would have affected atmospheric composition and surface temperature (Fanale *et al.*, 1992). In particular, injection of water and CO₂ into the atmosphere would have increased the surface temperature through greenhouse heating. On Mars, elevated temperatures would have led to further supplies of H₂O and CO₂ from polar caps and crustal reservoirs, which may have stabilized liquid water on the surface. For example, these processes could have accompanied volcanic activity associated with the Tharsis uprise (Phillips *et al.*, 2001).

On Venus, intensive volcanism ~0.5 Ga (Basilevsky *et al.*, 1997) would have been accompanied by IR-absorbing volcanic gases and a corresponding increase in surface temperature. Even a small fraction of water vapor in the products of volcanic degassing would have led to significant rising of atmospheric and lithospheric temperature. The temperature rise would have facilitated partial melting of igneous rocks, softening of the lithosphere, and trigger further volcanism (Solomon *et al.*, 1999). High-temperature reactions of atmospheric gases with hot surface materials could have led to significant alteration of surface rocks. In particular, degassed SO₂ and halogens could have been partially consumed by surface materials. Volcanically released H₂O would have been consumed by high-temperature oxidation of rocks (reactions [3] and [8]). It is possible that most chemical alteration currently observed on the surface occurred during the short but intense period of volcanic activity. Since the resurfacing event, abundances of minor atmospheric gases and the redox state may have been controlled by gas–mineral reactions. Changes in surface temperature altered the partial pressures of gases in gas–mineral-type reactions. For example, the oxidation state of the atmosphere and abundances of S-bearing gases could have been controlled by surface reactions that involve magnetite, hematite, and pyrite.

10.10.6 Summary and Unsolved Questions

Despite significant differences in environmental conditions on Mars and Venus, atmosphere–surface interactions reveal several common features. On both planets, atmospheric gases have oxidized surface materials. Dissociation of H₂O and hydrogen escape from early atmospheres could have caused partial oxidation of crustal materials. Although early atmospheric conditions could have favored formation of carbonates and hydrated minerals, current environments may not be feasible for carbonate formation. Throughout history, volcanic/impact degassing was followed by consumption of sulfur and halogens into surface materials through weathering reactions. On both planets, volcanic/impact degassing, surface temperature, and gas–mineral reactions are interrelated through the greenhouse heating and cloud feedbacks.

Specifics of atmosphere–surface interaction on Mars include: (1) seasonal cycles of condensation

and sublimation of H₂O and CO₂; (2) possible seasonal hydration and dehydration of salts; (3) hydration and dehydration of minerals during obliquity cycles; (4) low-temperature inhibition of thermochemical chemical reactions at the present surface; (5) a possible contribution of photo-stimulated gas–mineral-type reactions; (6) potential formation of the majority of secondary surface minerals by participation of aqueous solutions earlier in its history.

Venus' specifics includes (1) oxidation and sulfurization of surface rocks through gas–solid-type reactions; (2) isochemical weathering of individual solid phases with respect to nonvolatile elements at Venus' surface temperature (e.g., Al, Si, Mg, Fe, Ca, Na); (3) unlikely current hydration and a possibility of dehydration of early formed phases; (3) a strong altitudinal effect for chemistry and physics of gas–surface interactions; (4) possible condensation of heavy metal compounds on mountaintops; (5) the possibility of gas–mineral equilibria at the atmosphere–surface interface that determine the oxidation state, and sulfur and halogen content in the atmosphere; (6) the possibility for efficient atmosphere–surface reactions during intensive global volcanism.

A few of unsolved questions are the following:

- Under what environmental conditions were planetary surface materials oxidized and enriched in volatiles?
- What is the contribution of gas–mineral-type reactions to the composition of planetary surfaces?
- How have gas–mineral interactions changed over planetary history?
- What is the effect of irradiation on mineralogical transformations on the Martian surface?
- What are the abundances of C, halogens, and H in surface rocks on Venus?
- Do Venus' atmospheric gases equilibrate with each other and surface minerals?
- Which atmospheric gases are controlled by atmosphere–surface reactions on Venus?
- What are the alteration pathways and secondary mineralogies on Venus' lowlands and highlands?

Further progress could be made by integration of space missions, telescopic observations, experimental work, and theoretical models. A mission to Venus should aim to determine the composition of the near-surface atmosphere and the mineralogy of the surface. Experimental efforts focused on the kinetics of gas–mineral reactions on Venus and Mars should be complimented with coupled kinetic–thermodynamic models.

References

- Adamchik JA and Draper AL (1963) The temperature dependence of the Urey equilibrium and the problem of CO₂ content of the atmosphere of Venus. *Planetary and Space Science* 11: 1303–1307.
- Atreya SK, Wong A, Renno NO, et al. (2004) Oxidant enhancement in Martian dust devils and storms. *EOS Transactions of the American Geophysical Union* 85(47): Fall Meeting supplement abstract #SA13A-1123.
- Avduevsky VS, Godnev AG, Zakharov YV, et al. (1983) An estimate of the physical and mechanical characteristics of the soil of Venus from measurements of the impact overload during the landing of the Venera 13 and Venera 14 automatic interplanetary stations. *Cosmic Research* 21: 260–268.
- Baker VR, Komatsu G, Gulick VC, and Parker TJ (1997) Channels and valleys. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 757–796. Tucson, AZ: University of Arizona Press.
- Balme M and Greeley R (2006) Dust devils on Earth and Mars. *Reviews of Geophysics* 44: RG3003.
- Barin A, Clark BC, and Wanke H (1992) Surface chemistry and mineralogy. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 594–625. Tucson, AZ: University of Arizona Press.
- Barmin IV and Shevchenko AA (1983) Soil-scooping mechanism for the Venera 13 and Venera 14 unmanned interplanetary spacecraft. *Cosmic Research* 21: 118–121.
- Barsukov VL, Bazilevskii AT, Burba GA, Bobina NN, and Kriuchkov VP (1986) The geology and geomorphology of the Venus surface as revealed by the radar images obtained by Veneras 15 and 16. *Journal of Geophysical Research: Proceedings of the 16th LPSC* 91: D378–D398.
- Barsukov VL, Volkov VP, and Khodakovskiy IL (1982) The crust of Venus: Theoretical models of chemical and mineral composition. *Journal of Geophysical Research: Proceedings of the 13th LPSC* 87: A3–A9.
- Basilevsky AT, Head JW, and Abdrakhimov AM (2004) Impact crater air fall deposits on the surface of Venus: Areal distribution, estimated thickness, recognition in surface panoramas, and implications for provenance of sampled surface materials. *Journal of Geophysical Research* 109: E12003.
- Basilevsky AT, Head JW, Schaber GG, and Strom RG (1997) The resurfacing history of Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 1047–1086. Tucson, AZ: University of Arizona Press.
- Basilevsky AT, Kuzmin RO, Nikolaeva OV, et al. (1985) The surface of Venus as revealed by the Venera landings: Part II. *Geological Society of America Bulletin* 96: 137–144.
- Bell JF, McSween HY, Crisp JA, et al. (2000) Mineralogic and compositional properties of Martian soil and dust: Results from Mars Pathfinder. *Journal of Geophysical Research* 105: 1721–1755.
- Biemann K, Oro J, Toulmin P, et al. (1977) The search for organic substances and inorganic volatile compounds on the surface of Mars. *Journal of Geophysical Research* 82: 4641–4658.
- Booth MC and Kieffer HH (1978) Carbonate formation in Marslike environments. *Journal of Geophysical Research* 83: 1809–1815.
- Boynton WV, Feldman WC, Squyres SW, et al. (2002) Distribution of hydrogen in the near surface of Mars: Evidence for subsurface ice deposits. *Science* 297: 81–85.
- Bracken RA, Fegley B, and Arvidson RE (1995) Volatile transport on Venus and implications for surface geochemistry and geology. *Journal of Geophysical Research* 100: 1553–1563.

- Bridges JC, Catling DC, Saxton JM, Swindle TD, Lyon IC, and Grady MM (2001) Alteration assemblages in Martian meteorites: Implications for near-surface processes. *Space Science Reviews* 96(1/4): 365–392.
- Burt DR and Knauth LP (2002) Electrically conducting, Ca-rich brines, rather than water, expected in the Martian subsurface. *Journal of Geophysical Research* 108: 8026 (doi: 10.1029/2002JE001862).
- Campbell DB, Stacy NJS, Newman WI, et al. (1992) Magellan observations of extended impact-crater related features on the surface of Venus. *Journal of Geophysical Research* 97: 16249–16278.
- Carr MH (1996) *Water on Mars*, 229pp. New York: Oxford University Press.
- Catling DC and Moore J (2003) The nature of coarse-grained crystalline hematite and its implications for early environment of Mars. *Icarus* 165: 277–300.
- Christensen PR and Moore HJ (1992) The Martian surface layer. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 686–729. Tucson, AZ: University of Arizona Press.
- Christensen PR, Morris RV, Lane MD, Bandfield JL, and Malin MC (2001) Global mapping of Martian hematite deposits: Remnants of water-driven processes on early Mars. *Journal of Geophysical Research* 106: 23873–23886.
- Clancy RT, Sandor BJ, and Moriarty-Schieven GH (2004) A measurement of the 362 GHz absorption line of Mars atmosphere H₂O₂. *Icarus* 168: 116–121.
- Clark BC, Baird AK, Weldon RJ, Tsusaki DM, Schabel L, and Candelaria MP (1982) Chemical composition of Martian fines. *Journal of Geophysical Research* 87: 10059–10067.
- Clark BC, Kenley SL, O'Brien DL, Huss GR, Mack R, and Baird AK (1979) Heterogeneous phase reactions of Martian volatiles with putative regolith materials. *Journal of Molecular Evolution* 14: 91–102.
- Clark BC, Morris RV, McLennan SM, et al. (2005) Chemistry and mineralogy of outcrops at Meridiani Planum. *Earth and Planetary Science Letters* 240: 73–94.
- Clark BC and Van Hart DC (1981) The salts of Mars. *Icarus* 45: 370–378.
- Counselman CC, Gourevitch SA, King RW, Loriot GB, and Prinn RG (1979) Venus winds are zonal and retrograde below the clouds. *Science* 205: 85–87.
- Craddock RA and Howard AD (2002) The case for rainfall on a warm, wet early Mars. *Journal of Geophysical Research* 107(E11): 5111.
- Crumpler LS, Abeille JC, Senske DA, Keddie ST, Magee KP, and Head JW (1997) Volcanoes and centers of volcanism on Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 697–756. Tucson, AZ: University of Arizona Press.
- Donahue TM, Grinspoon DH, Hartle RE, and Hodges RR, Jr. (1997) Ion/neutral escape of hydrogen and deuterium: Evolution of water. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 385–414. Tucson, AZ: University of Arizona Press.
- Ekonovov AP, Golovin YuM, and Moshkin BE (1980) Visible radiation observed near the surface of Venus – Results and their interpretation. *Icarus* 41: 65–75.
- Encrenaz Th, Bézard B, Greathouse TK, et al. (2004) Hydrogen peroxide on Mars: Evidence for spatial and seasonal variations. *Icarus* 170: 424–429.
- Espenak F, Mumma MJ, Kostiuk T, and Zipoy D (1991) Ground-based infrared measurements of the global distribution of ozone in the atmosphere of Mars. *Icarus* 92: 252–262.
- Esposito LW, Bertaux J-L, Krasnopolsky V, Moroz VI, and Zasova LV (1997) Chemistry of lower atmosphere and clouds. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 415–458. Tucson, AZ: University of Arizona Press.
- Fanale FP, Postawko SE, Polack JB, Carr MH, and Pepin RO (1992) Mars: Epochal climate change and volatile history. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1135–1179. Tucson, AZ: University of Arizona Press.
- Farquhar J, Savarino J, Jackson TL, and Thiemens MH (2000) Evidence of atmospheric sulphur in the Martian regolith from sulphur isotopes in meteorites. *Nature* 404: 50–52.
- Fegley B and Treiman AH (1992) Chemistry of atmosphere–surface interactions on Venus and Mars. In: Luhmann JG, Tatryay M, and Pepin RO (eds.) *Venus and Mars: Atmospheres, Ionospheres, and Solar Wind Interactions*, Geophysical Monograph 66, pp. 7–71. Washington, DC: American Geophysical Union.
- Fegley B, Jr., Klingelhöfer G, Brackett RA, Izenberg N, Kremser DT, and Lodders K (1995) Basalt oxidation and the formation of hematite on the surface of Venus. *Icarus* 118: 373–383.
- Fegley B, Jr., Klingelhöfer G, Lodders K, and Widemann T (1997a) Geochemistry of surface–atmosphere interactions on Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 591–636. Tucson, AZ: University of Arizona Press.
- Fegley B, Jr. (2003) Venus. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 1, pp. 487–507. Oxford: Elsevier.
- Fegley B, Jr. and Prinn RG (1989) Estimation of the rate of volcanism on Venus from reaction rate measurements. *Nature* 337: 55–58.
- Fegley B, Jr., Zolotov MYu, and Lodders K (1997b) The oxidation state of the lower atmosphere and surface of Venus. *Icarus* 125: 416–439.
- Feldman WC, Prettyman TH, Maurice S, et al. (2004) Global distribution of near-surface hydrogen on Mars. *Journal of Geophysical Research* 109: E09006.
- Ferguson RL, Christensen PR, Bell JF, III, et al. (2006) Physical properties of the Mars Exploration Rover landing sites as inferred from Mini-TES-derived thermal inertia. *Journal of Geophysical Research* 111: E02S21.
- Florensky CP, Bazilevskii AT, Kriuchkov VP, et al. (1983) Venera 13 and Venera 14 – Sedimentary rocks on Venus? *Science* 221: 57–59.
- Florensky CP, Nikolaeva OV, Volkov VP, et al. (1983) Redox indicator ‘CONTRAST’ on the surface of Venus. *Lunar and Planetary Science XIV*: 203–204 (abstract, The Lunar and Planetary Institute, Houston).
- Florensky CP, Ronca LB, Basilevsky AT, et al. (1977) The surface of Venus as revealed by Soviet Venera 9 and 10. *Geological Society of America Bulletin* 88: 1537–1545.
- Ford PG and Pettengill GH (1983) Venus: Global surface radio emissivity. *Science* 220: 1379–1381.
- Frenkel MYa and Zabalaeva EV (1983) The solidification of effusive magmas on Venus and Earth. *Geochemistry International* 20(5): 43–47.
- Garvin JB, Head JW, Zuber MT, and Helfenstein P (1984) Venus: The nature of the surface from Venus panoramas. *Journal of Geophysical Research* 89: 3381–3399.
- Gellert R, Rieder R, Anderson RC, et al. (2004) Chemistry of rocks and soils in Gusev crater from the alfa particle X-ray spectrometer. *Science* 305: 829–832.
- Gellert R, Rieder R, Brückner J, et al. (2006) Alpha particle X-ray spectrometer (APXS): Results from Gusev crater and calibration report. *Journal of Geophysical Research* 111: E02S05.
- Gendrin A, Mangold N, Bibring J-P, et al. (2005) Sulfates in Martian layered terrains: The OMEGA/Mars Express view. *Science* 307: 1587–1591.
- Glotch TD, Bandfield JL, Christensen PR, et al. (2006) Mineralogy of the light-toned outcrop at Meridiani Planum as seen by the miniature thermal emission spectrometer and

- implications for its formation. *Journal of Geophysical Research* 111: E12S03.
- Goetz W, Bertelsen P, Binai CS, et al. (2005) Indication of drier periods on Mars from the chemistry and mineralogy of atmospheric dust. *Nature* 436: 62–65.
- Golombek MP, Grant JA, Crumpler LS, et al. (2006) Erosion rates at the Mars Exploration Rover landing sites and long-term climate change on Mars. *Journal of Geophysical Research* 111: E12S10.
- Gooding JL (1978) Chemical weathering on Mars. Thermodynamic stability of primary minerals (and their alteration products) from mafic igneous rock. *Icarus* 33: 483–513.
- Gooding JL (1992) Soil mineralogy and chemistry on Mars: Possible clues from salts and clays in SNC meteorites. *Icarus* 99: 28–41.
- Gooding JL, Arvidson R, and Zolotov MYu (1992) Physical and chemical weathering. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 626–651. Tucson, AZ: University of Arizona Press.
- Greeley R, Bender KC, Saunders RS, Schubert G, and Weitz CM (1997) Aeolian processes and features on Venus. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 547–590. Tucson, AZ: University of Arizona Press.
- Greeley R and Iversen JD (1985) *Wind as a Geological Process on Earth, Mars, Venus and Titan*, 333pp. Cambridge: Cambridge University Press.
- Greeley R, Lankaster N, Lee S, and Thomas P (1992) Martian aeolian processes, sediments, and features. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 730–766. Tucson, AZ: University of Arizona Press.
- Greeley R, Leach R, White B, Iversen J, and Pollack J (1980) Threshold windspeeds for sand on Mars: Wind tunnel simulations. *Geophysical Research Letters* 7: 121–124.
- Greeley R, Marshall JR, and Pollack JB (1987) Physical and chemical modification of the surface of Venus by windblown particles. *Nature* 327: 313–331.
- Grotzinger JP, Arvidson RE, Bell JF, et al. (2005) Stratigraphy and sedimentology of a dry to wet eolian depositional system, Burns formation, Meridiani Planum, Mars. *Earth and Planetary Science Letters* 240: 11–72.
- Gupta SC, Ahrens TJ, and Yang W (2001) Shock-induced vaporization of anhydrite and global cooling from the K/T impact. *Earth and Planetary Science Letters* 188: 399–412.
- Hashimoto GL and Abe Y (2001) Predictions of a simple cloud model for water vapor cloud albedo feedback on Venus. *Journal of Geophysical Research* 106: 14675–14690.
- Haskin LA, Wang A, Jolliff BL, et al. (2005) Water alteration of rocks and soils on Mars at the Spirit rover site in Gusev crater. *Nature* 436: 66–69.
- Head JW and Wilson L (1986) Volcanic processes and landforms on Venus: Theory predictions and observations. *Journal of Geophysical Research* 91: 9407–9446.
- Hoefen TM, Clark RN, Bandfield JL, Smith MD, Pearl JC, and Christensen PR (2003) Discovery of olivine in the Nili Fossae region of Mars. *Science* 302: 627–630.
- Huguenin RL (1973a) Photostimulated oxidation of magnetite 1. Kinetics and alteration phase identification. *Journal of Geophysical Research* 78: 8281–8493.
- Huguenin RL (1973b) Photostimulated oxidation of magnetite 2. Mechanism. *Journal of Geophysical Research* 78: 8495–8506.
- Huguenin RL (1974) The formation of goethite and hydrated clay minerals on Mars. *Journal of Geophysical Research* 79: 3895–3905.
- Huguenin RL (1976) Mars: Chemical weathering as a massive volatile sink. *Icarus* 28: 203–212.
- Huguenin RL (1982) Chemical weathering and the Viking biology experiments on Mars. *Journal of Geophysical Research* 87: 10069–10082.
- Huguenin RL, Miller KJ, and Harwood WS (1979) Frost weathering on Mars: Experimental evidence for peroxide formation. *Journal of Molecular Evolution* 14: 103–132.
- Jakosky BM and Haberle RM (1992) The seasonal behavior of water on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 969–1016. Tucson, AZ: University of Arizona Press.
- James PB, Kieffer HH, and Paige DA (1992) The seasonal cycle of carbon dioxide on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 934–968. Tucson, AZ: University of Arizona Press.
- Johnson NM and Fegley B (2003) Tremolite decomposition on Venus II. Products, kinetics, and mechanism. *Icarus* 164: 317–333.
- Johnson NM and Fegley B (2005) Phlogopite decomposition, water, and Venus. *36th Lunar and Planetary Science Conference* (CD-ROM), abstract no 1992. League City, TX: Lunar and Planetary Science.
- Kasting JF (1991) CO₂ condensation and the climate of early Mars. *Icarus* 94: 1–13.
- Kemurdzhian AL, Brodskii PN, Gromov VV, et al. (1983) Preliminary results of determining the physical and mechanical properties of the soil of Venus by the Soviet automatic stations Venera 13 and Venera 14. *Cosmic Research* 21: 253–259.
- Kieffer HH, Jakosky BM, and Snyder CW (1992) The planet Mars: From antiquity to the present. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1–33. Tucson, AZ: University of Arizona Press.
- Kieffer HH and Zent AP (1992) Quasi-periodic climate change on Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 1180–1218. Tucson, AZ: University of Arizona Press.
- Klein HP (1978) The Viking biological experiments on Mars. *Icarus* 34: 666–674.
- Klingelhöfer G, Morris RV, Bernhardt B, et al. (2004) Jarosite and hematite at Meridiani Planum from Opportunity's Mössbauer spectrometer. *Science* 306: 1740–1745.
- Klose KB, Wood JA, and Hashimoto A (1992) Mineral equilibria and the high radar reflectivity of Venus mountaintops. *Journal of Geophysical Research* 97: 16353–16369.
- Knauth LP, Burt DM, and Wohletz KH (2005) Impact origin of sediments at the Opportunity landing site on Mars. *Nature* 438: 1123–1128.
- Krasnopolsky VA and Feldman PD (2001) Detection of molecular hydrogen in the atmosphere of Mars. *Science* 294: 1914–1917.
- Krasnopolsky VA, Maillard JP, and Owen TC (2004) Detection of methane in the Martian atmosphere: Evidence for life? *Icarus* 172: 537–547.
- Krasnopolsky VA and Pollack JB (1994) H₂O–H₂SO₄ system in Venus' clouds and OCS, CO, and H₂SO₄ profiles in Venus' troposphere. *Icarus* 109: 58–78.
- Ksanfomality LV, Gorshkova NV, and Khondyrev VK (1993) Wind velocity near the surface of Venus from acoustic measurements. *Cosmic Research* 21: 161–167.
- Kuzmin RO, Christensen PR, and Zolotov MYu (2004) Global mapping of Martian bound water at 6.1 microns based on TES data: Seasonal hydration–dehydration of surface minerals. *Lunar and Planetary Science* XXXV: abstract no 1810 (CD-ROM).
- Langevin Y, Poulet F, Bibring J-P, and Gondet B (2005) Sulfates in the North polar region of Mars detected by OMEGA/Mars express. *Science* 307: 1584–1586.
- Lewis JS (1970) Venus: Atmospheric and lithospheric composition. *Earth and Planetary Science Letters* 10: 73–80.
- Lewis JS and Prinn RG (1984) *Planets and Their Atmospheres*, 470 pp. Orlando, FL: Academic Press.
- McElroy MB, Kong TY, and Yung YL (1977) Photochemistry and evolution of Mars' atmosphere: A Viking perspective. *Journal of Geophysical Research* 82: 4379–4388.

- McSween HY, Jr. (2003) Mars. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry*, vol. 1, pp. 601–621. Oxford: Elsevier.
- McSween HY, Wyatt MB, Gellert R, et al. (2006) Characterization and petrologic interpretation of olivine-rich basalts at Gusev crater, Mars. *Journal of Geophysical Research* 111(E2): E02S10.
- Mellon MT and Jakosky BM (1995) The distribution and behavior of Martian ground ice during past and present epochs. *Journal of Geophysical Research* 100: 11781–11799.
- Michalski JR, Kraft MD, Sharp TG, Williams LB, and Christensen PR (2005) Mineralogical constraints on the high-silica Martian surface component observed by TES. *Icarus* 174: 161–177.
- Ming DW, Mittlefehldt DW, Morris RV, et al. (2006) Geochemical and mineralogical indicators for aqueous processes in the Columbia Hills of Gusev crater, Mars. *Journal of Geophysical Research* 111: E02S12.
- Mitrofanov IG, Zuber MT, Litvak ML, et al. (2004) CO₂ snow depth and subsurface water–ice abundance in the Northern hemisphere of Mars. *Science* 300: 2081–2084.
- Morris RV, Klingelhöfer G, Schröder C, et al. (2006) Mössbauer mineralogy of rock, soil, and dust at Gusev crater, Mars: Spirit's journey through weakly altered olivine basalt on the plains and pervasively altered basalt in the Columbia Hills. *Journal of Geophysical Research* 111: E02S13.
- Morris RV and Lauer HV, Jr. (1980) The case against UV photostimulated oxidation of magnetite. *Geophysical Research Letters* 7: 605–608.
- Morris RV and Lauer HV, Jr. (1981) Stability of goethite (α -FeOOH) and lepidocrocite (γ -FeOOH) to dehydration by UV radiation – Implications for their occurrence on the Martian surface. *Journal of Geophysical Research* 86: 10893–10899.
- Morris RV, Shetler TD, Scheinost AC, et al. (2000) Mineralogy, composition, and alteration of Mars Pathfinder rocks and soil: Evidence from multispectral, elemental, and magnetic data on terrestrial analogues, SNC meteorite, and Pathfinder samples. *Journal of Geophysical Research* 105: 1757–1817.
- Mueller RF (1964) A chemical model for the lower atmosphere of Venus. *Icarus* 3: 285–298.
- Mueller RF (1968) Sources of HCl and HF in the atmosphere of Venus. *Nature* 220: 55–57.
- Mukhin LM, Koscheev AP, Dikov YuP, Huth J, and Waenke H (1996) Experimental simulations of the photodecomposition of carbonates and sulphates on Mars. *Nature* 379: 141–143.
- Mustard JF, Poulet F, Gendrin A, et al. (2005) Olivine and pyroxene diversity in the crust of Mars. *Science* 307: 1594–1597.
- Nordstrom DK, Alpers CN, Ptacek CJ, and Blowes DW (2000) Negative pH and extremely acidic mine waters from Iron Mountain, California. *Environmental Science and Technology* 34: 254–258.
- O'Connor JT (1968) Mineral stability at the Martian surface. *Journal of Geophysical Research* 73: 5301–5311.
- Owen T (1992) The composition and early history of the atmosphere of Mars. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 818–834. Tucson, AZ: University of Arizona Press.
- Oyama VI and Berdahl BJ (1977) The Viking gas exchange experiment results from Chryse and Utopia surface samples. *Journal of Geophysical Research* 82: 4669–4676.
- Oyama VI and Berdahl BJ (1979) A model of Martian surface chemistry. *Journal of Molecular Evolution* 14: 199–210.
- Parkinson TD and Hunten DM (1972) Spectroscopy and aeronomy of O₂ on Mars. *Journal of the Atmospheric Sciences* 29: 1380–1390.
- Pettengill GH, Campbell BA, Campbell DB, and Simpson RA (1997) Surface scattering and dialectical properties. In: Bouger SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 527–546. Tucson, AZ: University of Arizona Press.
- Phillips RJ, Zuber MT, Solomon SC, et al. (2001) Ancient geo-dynamics and global-scale hydrology on Mars. *Science* 291: 2587–2591.
- Pieters CM, Head JW, Patterson W, et al. (1986) The color of the surface of Venus. *Science* 234: 1379–1383.
- Pollack JB, Kasting JF, Richardson SM, and Poliakoff K (1987) The case for a wet, warm climate on early Mars. *Icarus* 71: 203–224.
- Pope KO, Baines KH, Ocampo AC, and Ivanov BA (1997) Energy, volatile production, and climatic effects of the Chicxulub Cretaceous/Tertiary impact. *Journal of Geophysical Research* 102: 21645–21664.
- Poulet F, Bibring J-P, Mustard JF, et al. (2005) Phyllosilicates on Mars and implications for early Martian climate. *Nature* 438: 623–627.
- Prinn RG and Fegley B, Jr. (1987) Bolide impacts, acid rain, and biospheric traumas at the Cretaceous–Tertiary boundary. *Earth and Planetary Science Letters* 83: 1–15.
- Rieder R, Economou T, Wanke H, et al. (1997) The chemical composition of Martian soil and rocks returned by the mobile alpha proton X-ray spectrometer: Preliminary results from the X-ray mode. *Science* 278: 1771–1774.
- Rieder R, Gellert R, Anderson RC, et al. (2004) Chemistry of rocks and soils at Meridiani Planum from the alpha particle X-ray spectrometer. *Science* 306: 746–749.
- Ronov AB and Yaroshevsky AA (1976) A new model for the chemical structure of the Earth crust. *Geochemistry International* 13(6): 89–121.
- Schaefer L and Fegley B, Jr. (2004) Heavy metal frost on Venus. *Icarus* 168: 215–219.
- Segura TL, Toon OB, Colaprete A, and Zahnle K (2002) Environmental effects of large impacts on Mars. *Science* 298: 1977–1980.
- Settle M (1979) Formation and deposition of volcanic sulfate aerosols. *Journal of Geophysical Research* 84: 8343–8354.
- Shimazaki T and Shimizu M (1979) The seasonal variation of ozone density of the Martian atmosphere. *Journal of Geophysical Research* 84: 1269–1276.
- Shkuratov YuG, Kreslavsky MA, and Nikolayeva OV (1987) Albedo-color diagram of the Venusian surface and its interpretation. *Solar System Research* 21(2): 94–102.
- Soderblom LA (1992) The composition and mineralogy of the Martian surface from spectroscopic observations. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 557–593. Tucson, AZ: University of Arizona Press.
- Solomon SC, Bullock MA, and Grinspoon DH (1999) Climate change as a regulator of tectonics on Venus. *Science* 286: 87–90.
- Squyres SW, Knoll AH, Arvidson RE, et al. (2006) Two years at Meridiani Planum: Results from the Opportunity rover. *Science* 313: 1403–1407.
- Surkov YuA, Barsukov VL, Moskalyeva LP, Kharyukova VP, and Kemurdzhian AL (1984) New data on the composition, structure, and properties of Venus rock obtained by Venera-13 and Venera-14. *Journal of Geophysical Research: Proceedings of the 15th LPSC* 89: 393–402.
- Surkov YuA, Kirnozov FF, Glazov VN, Dunchenko AG, Tatsy LP, and Soborov OP (1987) Uranium, thorium, and potassium in the Venusian rocks at the landing sites of Vega-1 and Vega-2. *Journal of Geophysical Research: Proceedings of the 17th LPSC* 92: 537–540.
- Surkov YuA, Moskalyova LP, Kharyukova VP, Dudin AD, Smirnov GG, and Zaitseva SYe (1986) Venus rock composition at the Vega-2 landing site. *Journal of Geophysical Research: Proceedings of the 16th LPSC* 91: 215–218.

- Symonds RB, Reed MH, and Rose WI (1992) Origin, speciation, and fluxes of trace-element gases at Augustine volcano, Alaska: Insights into magma degassing and fumarolic processes. *Geochimica et Cosmochimica Acta* 56: 633–657.
- Symonds RB, Rose WI, Bouth GJS, and Gerlach TM (1994) Volcanic-gas studies: Methods, results, and applications. In: Carroll MR and Holloway JR (eds.) *Volatiles in Magmas*, pp. 1–66. Washington, DC: Mineralogical Society of America.
- Tanner SB, Kerrick DM, and Lasaga AC (1985) Experimental kinetic study of the reaction: Calcite + quartz → wollastonite + carbon dioxide, from 1 to 3 kilobars and 500° to 850°C. *American Journal of Science* 285: 577–620.
- Terradellas J and Bonnetaud L (1973) Nature of the chemical reactions during the reaction of sulfur dioxide with calcium carbonate. *Bulletin de la Societe Chimique de France* 6: 1903–1908.
- Toon OB, Zahnle K, Morrison D, Turco RP, and Covey C (1997) Environmental perturbations caused by the impacts of asteroids and comets. *Reviews of Geophysics* 35: 41–78.
- Volkov VP (1992) Volatiles in atmosphere and crust. In: Barsukov VL, Basilevsky AT, Volkov VP, and Zharkov VN (eds.) *Venus Geology, Geochemistry, and Geophysics*, pp. 200–210. Tucson, AZ: University of Arizona Press.
- Volkov VP, Zolotov MYu, and Khodakovskiy IL (1986) Lithospheric–atmospheric interaction on Venus. In: Saxena SK (ed.) *Chemistry and Physics of Terrestrial Planets*, pp. 136–190. New York: Springer.
- Von Zahn U, Kumar S, Niemann H, and Prinn R (1983) Composition of the Venus atmosphere. In: Hunten DM, Colin L, Donahue TM, and Moroz VM (eds.) *Venus*, pp. 299–430. Tucson, AZ: University of Arizona Press.
- Wänke H and Dreibus G (1987) Volatiles on Earth and Mars: A comparison. *Icarus* 71: 225–240.
- Wood JA (1997) Rock weathering on the surface of Venus. In: Bougher SW, Hunten DM, and Phillips RJ (eds.) *Venus II*, pp. 637–666. Tucson, AZ: The University of Arizona Press.
- Urey HC (1952) *The Planets*. New Haven, UK: Yale University Press.
- Yen AS, Gellert R, Schröder C, et al. (2005) An integrated view of the chemistry and mineralogy of Martian soils. *Nature* 436: 49–54.
- Yen AS, Kim SS, Hecht MH, Frant MS, and Murray B (2000) Evidence that the reactivity of the Martian soil is due to superoxide ions. *Science* 289: 1909–1912.
- Yen AS, Murray B, Rossman GR, and Grunthaner FJ (1999) Stability of hydroxylated minerals on Mars: A study on the effects of exposure to ultraviolet radiation. *Journal of Geophysical Research* 104: 27031–27041.
- Yung YL and DeMore WD (1999) *Photochemistry of Planetary Atmospheres*, 456 pp. New York: Oxford University Press.
- Zent AP and McKay CP (1994) The chemical reactivity of the Martian soil and implications for future missions. *Icarus* 108: 146–157.
- Zolotov MYu (1996) A model for the thermal equilibrium of the surface venusian atmosphere. *Geochemistry International* 33(10): 80–100.
- Zolotov MYu, Fegley B, Jr., and Lodders K (1997) Hydrous silicates and water on Venus. *Icarus* 130: 475–494.
- Zolotov MYu and Mironenko MV (2007) Timing of acid weathering on Mars: A kinetic-thermodynamic assessment. *Journal of Geophysical Research* 111 (in press).
- Zolotov MYu and Volkov VP (1992) Chemical processes on the planetary surface. In: Barsukov VL, Basilevsky AT, Volkov VP, and Zharkov VN (eds.) *Venus Geology, Geochemistry, and Geophysics*, pp. 177–200. Tucson, AZ: University of Arizona Press.

10.11 Water on the Terrestrial Planets

J. Helbert and E. Hauber, Institute for Planetary Research, DLR, Berlin, Germany

D. Reiss, Institut für Planetologie, Westfälische Wilhelms-Universität Münster, Münster, Germany

© 2007 Elsevier B.V. All rights reserved.

10.11.1	Introduction	372
10.11.1.1	Geophysical Constraints	372
10.11.1.1.1	Stability of water and ice	372
10.11.1.1.2	Adsorption	374
10.11.2	Observational Evidence	375
10.11.2.1	Spectroscopy, Morphology, and Radar	375
10.11.2.1.1	Visual and infrared spectroscopic features of ice and bound water	375
10.11.2.1.2	Gamma-ray and neutron spectroscopy	376
10.11.2.1.3	Morphology	377
10.11.2.1.4	Radar	378
10.11.3	Water on the Surface	378
10.11.3.1	Valley Networks	379
10.11.3.1.1	Morphology and formation	379
10.11.3.1.2	Ages	380
10.11.3.1.3	Source of water and climate	380
10.11.3.2	Sedimentary Processes	381
10.11.3.2.1	The orbiter view	381
10.11.3.2.2	The view of the MER rovers	382
10.11.3.3	Outflow Channels	382
10.11.3.3.1	Morphology	382
10.11.3.3.2	Morphometry/discharge	384
10.11.3.3.3	Formation	385
10.11.3.4	Oceans	385
10.11.3.4.1	Models involving oceans	385
10.11.3.4.2	Assessment of geomorphic evidence	386
10.11.3.4.3	Tests of ocean hypotheses	387
10.11.3.5	Glacial, Permafrost, and Ground-Ice Features	388
10.11.3.5.1	Glacial landforms	388
10.11.3.5.2	Permafrost and ground-ice landforms	389
10.11.3.6	Polar Caps	393
10.11.3.7	Large Debris Flows	394
10.11.3.8	Gullies	395
10.11.3.8.1	Morphology	395
10.11.3.8.2	Distribution	396
10.11.3.8.3	Formation	396
10.11.3.8.4	Age	397
10.11.4	Water in Mantle and Crust	397
10.11.4.1	Regolith	398
10.11.4.1.1	Moon and Mercury	398
10.11.4.1.2	Mars	399
10.11.4.2	Mineralogy	400
10.11.4.2.1	Methods	401
10.11.4.2.2	Basalt and dust	401
10.11.4.2.3	Hematite	401

10.11.4.2.4	Phyllosilicates	402
10.11.4.2.5	Sulfates	402
10.11.4.2.6	Carbonates	403
10.11.4.2.7	Zeolites	403
10.11.4.3	Geophysical Consequences	403
10.11.5	Evolution of Water and Climate	405
10.11.6	Summary and Outlook	407
References		408

10.11.1 Introduction

Water is the second most abundant molecule in our solar system, following only molecular hydrogen. It plays an important role in many planetary processes and has significant importance for the habitability of a planetary body. Despite the importance of water its complex physical behavior is still only partly understood.

This chapter provides an overview of the current knowledge of water and water-related processes on the terrestrial planets from a geophysical as well as a morphological point of view. Naturally Mars plays a featured role here as it is the only terrestrial planet where we have clear evidence for present and past widespread existence, and activity of water on and close to the surface. In addition radar observations indicate the presence of water-ice deposits on Mercury inside permanently shadowed craters close to the poles of the planet. As the focus of this chapter is on terrestrial planets, we will not discuss the icy moons of the giant planets which will be dealt with in Chapter 10.13 or Pluto which is discussed in Chapter 10.16.

This chapter is organized into six sections. Section 10.11.1 presents a general discussion of the geophysical constraints for the stability of ice and water focussing on the conditions found on the terrestrial planets. This is followed by a discussion of observational methods to detect the presence of water and ice. We then discuss water and ice on the terrestrial planets starting with water on the surface of the terrestrial planets including a discussion of its morphological impact. From there we move on to discuss water in the mantle and crust of terrestrial planets and discuss briefly the effects of water in the interior of terrestrial planets in Section 10.11.4. This topic is also covered in Chapter 10.09. Finally a brief discussion on the evolution of water and climate on the terrestrial planets is included. For a more detailed discussion of the interactions between atmospheres and surfaces of terrestrial planets we refer to

Chapter 10.10. We conclude the chapter with a summary and outlook in Section 10.11.6. This includes a very brief discussion of necessary future measurements. For a more detailed discussion of planetary exploration strategies we refer to Chapter 10.18.

All terrestrial planets are topic of this chapter, but in many discussions Mars will take centerstage as it is the planet showing the most striking evidence of water activity in the past and even in the present.

10.11.1.1 Geophysical Constraints

10.11.1.1.1 Stability of water and ice

Water exists as a liquid at pressures above 611 hPa and temperatures above 273 K, the triple point of water. Under these conditions the solid, liquid, and vapor phases coexist. Branching from the triple point are the phase-transition curves for solid–liquid, liquid–vapor, and solid–vapor phase changes. Depending on the pressure, water can stay in liquid state up to more than 600 K. The phase diagram of water (Figure 1) reflects the complex behavior of the substance.

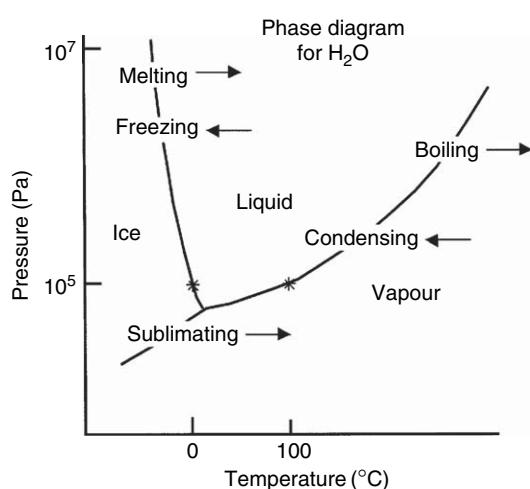


Figure 1 Phase diagram of water.

On planetary surfaces, temperatures are determined by the solar irradiance, which is a function of the solar luminosity (L) and the distance (D) as L/D^2 . Thus, phase conditions for liquid water are dependent on the distance from the central star. Conservative estimates for distances at which water can be stable as a liquid in our solar system are between 0.95 and 1.37 AU (Kasting *et al.*, 1993). Thus of the terrestrial planets only Mars (apart from Earth) has the potential for transient liquid water on the surface as its orbit is just at the outer edge of the estimate by Kasting *et al.* (1993). For the Moon, which is inside the distance limits but has no atmosphere, and for Mercury only ice might be stable on or close to the surface within permanently shadowed regions. The present-day surface of Venus is too hot to support liquid water.

On Mars, with its thin atmosphere, liquid water might be a transient occurrence. If the total pressure of atmospheric gas exceeds 610 Pa, pure liquid water may be stable against boiling, but it is still unstable with respect to evaporation. The boiling point of a liquid is that temperature at which its saturation vapor pressure equals the total external pressure. This total pressure can be supplied in the case of Mars, for example, by CO₂ and does not refer to the partial pressure of water vapor. For example, at 283 K the saturation vapor pressure of liquid water is \sim 12 mbar. With these atmospheric conditions liquid water would be stable against boiling even if the partial pressure of water in the atmosphere is less than the saturation vapor. Since the atmospheres of Mars and Earth are not saturated, liquid water is not stable on the surface of either planet with respect to evaporation. Stability of liquid water on the surface means therefore in general that liquid water, if it could form, would be stable against boiling or freezing but (most likely) unstable to evaporation. For more details see the discussion by Haberle *et al.* (2001).

As pointed out by Richardson and Mischna (2005), boiling is only of major importance as a process because it leads to a much more rapid conversion of liquid to vapor than evaporation. Bubble development in the interior of the liquid is allowed and dynamically favored, if the saturation vapor pressure of the liquid exceeds the total atmospheric pressure. It enables the vaporization of the liquid to occur throughout the full volume of liquid. The rough surface created by boiling might also lead to a more effective heat exchange. In general it can be summarized that liquid water might exist if surface pressures

and temperatures fall between the melting point of ice and the boiling point of liquid water.

For an atmosphereless body like Mercury or the Moon the stability of ice is determined purely by the sublimation rate as discussed recently by Andreas (2007). The standard equation for the sublimation (or evaporation) of ice (or pure water) in vacuum is given by Estermann (1955) as

$$S = p_{\text{sat},i}(T) \left[\frac{M_w}{2\pi RT} \right]^{1/2}$$

where S is a mass flux in kg m⁻² s⁻¹ and $p_{\text{sat},i}(T)$ is the saturation pressure in hPa for a ice surface at temperature T (in Kelvin). M_w is the molecular weight of water, and R the universal gas constant. Using the sublimation equation it is straightforward to derive the temperature-dependent sublimation if the saturation vapour pressure curve is well known especially at low temperatures.

Murphy and Koop (2005) provided recently a new function for the temperature dependence

$$p_{\text{sat}}(T) = 0.01 \exp \left(9.550426 - \frac{5723.265}{T} + 3.3068 \ln T - 0.00728332 \right)$$

The expression is valid for $T=110$ –273.15 K. $p_{\text{sat}}(T)$ is in hPa, and T in K.

For a planet like Mars, the sublimation can be limited by the saturation of the atmosphere with water vapor. The saturation can be expressed in terms of relative humidity

$$r_b = \frac{\rho \cdot R_v \cdot T}{p_{\text{sat}}(T)}$$

where ρ is the water vapor density, T the temperature, p_{sat} the saturation pressure, and $R_v = 461.5 \text{ J K}^{-1} \text{ kg}^{-1}$.

For Mars the abundance of water vapor in the atmosphere pr is often given in precipitable microns, describing the amount of water if the whole atmospheric column would be condensed on the surface.

Assuming the simple case that water follows a scale height-dependent distribution this can be converted to water vapor density using

$$\rho(t) = \frac{\rho_{\text{H}_2\text{O}} \cdot pr}{H(t) \cdot 100}$$

where $\rho_{\text{H}_2\text{O}}$ is the density of water, pr the water abundance in precipitable microns, and $H(t)$ the scale height which can vary as a function of

Martian seasons. One has to be cautious, however, because the vertical distribution of water in the Martian atmosphere is not readily describable only by a scale height. There are many instances where water does not follow a scale height, or where there is an inversion and the number density increases with altitude.

Salinity does not influence the solid–vapor transition temperature, as salts are not incorporated into ice. It influences however the solid–liquid transition temperature. Addition of salt to water causes the triple point to migrate to lower temperature and pressures (see Figure 2). Liquid water may exist in the form of salt saturated solution (also called brines) at temperatures significantly below 273 K. Brines with a high content of CaCl_2 remain liquid up to 220–225 K (see, e.g., Brass, 1980; Kuzmin and Zabalueva, 1998). Haberle *et al.* (2001) have pointed out that while pure water could exist over 29% of the surface of Mars for 37 sols each year, a NaCl eutectic solution could remain liquid over most of the planet at some time of the year and up to 100 sols in some regions. As discussed before this refers only to transient liquid water, as liquid water is stable nowhere on the Martian surface. The formation of eutectic solutions of CaCl_2 , with its much lower eutectic temperature, would greatly increase the extent and duration of liquid on Mars. Sears and Chittenden (2005) showed recently that brine formation will increase the stability of water on Mars not only by extending the liquid temperature range, but also by considerably decreasing the evaporation rate. At the eutectic temperature of NaCl brine, evaporation rates are ~ 30 times less than pure water; at the

eutectic temperatures of CaCl_2 evaporation rates are slightly lower. In longer contact with the atmosphere the water will still evaporate leaving behind evaporites. Such deposits have been suggested based on the geochemical studies by the Mars Exploration Rovers (MERs) (Squyres *et al.*, 2003; Christensen *et al.*, 2003). For a short time, brines could appear on the Martian surface or nearby even under present-day conditions and would be stable over longer time spans than pure liquid water.

10.11.1.2 Adsorption

Hydrogen bonds can form not only between water molecules, but also between a water molecule and a silicate surface. The water molecules are typically bound in monolayers to the grain surface. There are a number of theories describing the amount of water that a certain material can hold. Following Jakosky (1986) with σ as the number of molecules bonded we get

$$\sigma = \frac{Ap(1-\alpha)}{(2\pi MRT)^{1/2}} \tau_0 e^{Q/RT}$$

In this equation α is the fraction of molecules striking the surface A which do not stick and it is assumed that the vibrating water molecules have a Maxwellian energy distribution. Q is the binding energy, R the gas constant, M the molecular weight, and T the temperature.

Defining $\theta = \sigma/\sigma_0$ as the fraction of monolayer of molecules present yields

$$\theta = \frac{kp}{1 + kp}$$

where

$$k = \frac{A}{(2\pi MRT)^{1/2}} \frac{\tau}{\sigma_0} (1 - \alpha)$$

These two equations form the Langmuir adsorption isotherm (Langmuir, 1915). This is the most simple approach because it limits adsorption to one monolayer. For most real materials it fails, because they can adsorb more than one monolayer. A more realistic approach was proposed by Brunauer *et al.* (1938), termed the Brunauer–Emmett–Teller (BET) theory. Using a similar approach as discussed above the isotherm can be expressed as

$$\theta = \frac{kx}{(1-x)(1-x+kx)}$$

where $x = P/P_s$ and P_s is the saturation pressure of the molecule. As adsorbed water molecules behave

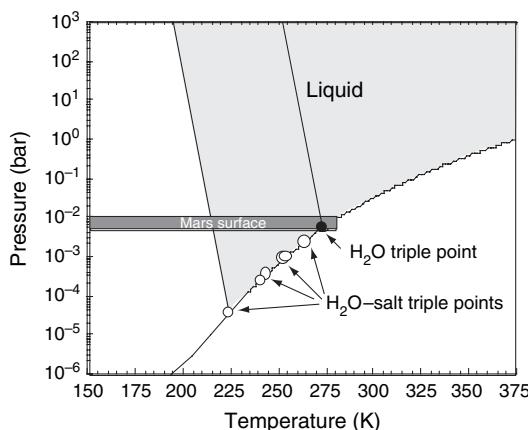


Figure 2 Phase diagram for a $\text{H}_2\text{O}-\text{CaCl}_2$ brine (Bodnar, 2001).

typically more like a liquid than a solid (quasi-liquid), P_s is the saturation pressure with respect to the liquid.

These rather simplified theories of adsorption do not take into account a number of parameters, including grain-size distribution, ambient pressure, and temperature. Applying these theories to Mars, for example, has shown the need for measurements of the adsorption isotherms of real materials and realistic planetary conditions. Fanale and Cannon (1971, 1974) measured the adsorption of H_2O on basalt for $T = 250\text{--}300\text{ K}$ and $P = 2\text{--}2000\text{ Pa}$. They found that approximately 10^{-3} kg of water per kg were adsorbed. To extrapolate their measurement to Mars condition usually a least square fit of an empirical isotherm is used (Zent and Quinn, 1997)

$$\rho = \rho_s A_s \frac{\gamma P^{0.51}}{e^{\delta/T}}$$

This equation gives the adsorbed load of water per cubic meter of regolith. Here ρ_s is the density of the regolith, A_s the specific surface area, $\gamma = 2.0487 \cdot 10^{-12} \text{ kg m}^{-2} \text{ Pa}^{-1}$ and $\delta = -2679.8\text{ K}$.

Using this isotherm and assuming a specific surface area of $17\text{ m}^2 \text{ g}^{-1}$ (Ballou *et al.*, 1978) it was predicted that the Martian regolith can hold about 2.1 kg m^{-3} of water and that formation of ground ice would be suppressed by this effective adsorption process (Flasar and Goody, 1976; Toon *et al.*, 1980).

Zent and Quinn (1995) made the first measurements of adsorption of water onto palagonite and montmorillonite at conditions appropriate to Mars. The palagonite data fitted best to a Toth isotherm

$$\theta = K^* P / [1 + (K^* P)^v]^{1/v}$$

and the montmorillonite data to a Freundlich isotherm

$$\theta = (K^* P)^v / [1 + (K^* P)]^v$$

with $K^* = K_0 e^{\varepsilon/T}$ for both isotherms. For palagonite, the fitting parameters are $K_0 = 7.54 \times 10^{-9}$, $\varepsilon = 2697.2$ and $v = 0.4734$ and for montmorillonite the fitting parameters are $K_0 = 2.207 \times 10^{-9}$, $\varepsilon = 4157.2$, and $v = 0.6593$.

With these new isotherms acquired under realistic conditions Zent and Quinn (1995) found that adsorption is about an order of magnitude less efficient than previously thought and formation of ground ice is not suppressed by adsorption. Zent *et al.* (2001) extended these studies to clays, namely smectites. They discovered that they play no significant role for diurnal

cycles of water, because the equilibration time is much longer than a Martian day and the abundances of adsorbed water are much smaller than predicted.

In a recent study Jänenchen *et al.* (2006) showed that microporous minerals (as nontronite, montmorillonite, chabazite, and clinoptilolite) might remain hydrated under present Martian atmospheric conditions and hold up to 2.5–25 wt.% of water in their void volumes at a partial water vapor pressure of 0.001 mbar in a temperature range of 333–193 K. These results are in contradiction to the much lower values reported by Zent *et al.* (1997, 2001). It should be however be noted that the measurements by Jänenchen *et al.* (2006) have been performed at 257–333 K and have only been extrapolated to the Martian temperature range. This extrapolation might be a reason for the disagreement, and stresses the importance of measurements under realistic Martian conditions.

10.11.2 Observational Evidence

10.11.2.1 Spectroscopy, Morphology, and Radar

10.11.2.1.1 Visual and infrared spectroscopic features of ice and bound water

H_2O is a bent polar molecule and has as such three fundamental vibration modes – a symmetric stretch mode at $2.738\text{ }\mu\text{m}$ (v_1), a bending mode at $6.270\text{ }\mu\text{m}$ (v_2), and an asymmetric stretch mode at $2.663\text{ }\mu\text{m}$ (v_3). In the vapor phase all absorptions predicted by group theory are observed. Phase changes into the liquid and solid phase change the spectral properties of water. In general, hydrogen bonds tend to broaden and intensify the O–H stretching features and shift them to longer wavelengths (Hamilton and Ibers, 1968). In liquid water, the fundamental vibration modes shift to $v_1 = 3.106\text{ }\mu\text{m}$, $v_2 = 6.079\text{ }\mu\text{m}$, and $v_3 = 2.903\text{ }\mu\text{m}$ (Clark, 1999).

Water frost and ice are generally bright and spectrally flat in the visual wavelength range. They show various diagnostic features in the ultraviolet (UV) and in the near- and mid-infrared (see Figure 3). Water ice has fundamental vibration modes at 3.2 and $6.1\text{ }\mu\text{m}$ and exhibits strong combination overtone bands centered near 1.5 and $2.0\text{ }\mu\text{m}$. The fact that these bands are so strong makes them prone to saturation. Therefore, the weaker overtones observed at 1.26 , $1.04\text{ }\mu\text{m}$, and shorter wavelengths can often provide more useful information (Clark, 1981; Roush *et al.*, 1993). The unsaturated ice

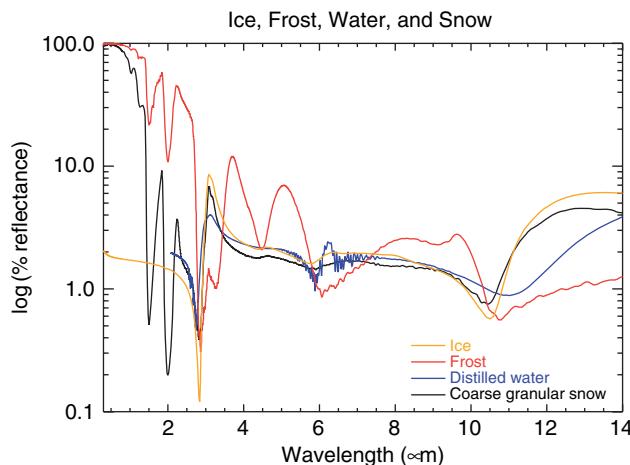


Figure 3 Measured reflectance spectra of different types of water deposits on planetary surface – water ice, frost, distilled water, and coarse granular snow. Data from the ASTER Spectral Library through the courtesy of the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, 1999, California Institute of Technology.

bands have distinctive positions, widths, and shapes different from the bands due to molecular water and OH⁻ absorptions on minerals as discussed below.

Many minerals can incorporate H₂O and/or OH⁻ in their structure. H₂O can also be absorbed on mineral surfaces, located in interlayer sites of smectite clays or trapped as fluid inclusions. All these forms are commonly found in terrestrial rocks as a result of weathering, hydrothermal alteration, or formed by sedimentary precipitation. The near-infrared region is ideally suited to search for water- or hydroxyl-bearing minerals (Farmer, 1974; Salisbury *et al.*, 1991; Roush *et al.*, 1993). Dominant in this region is a strong absorption feature around 3 μm. This feature is caused by an overlap of the symmetric and antisymmetric stretching modes with the first overtone of the strong 6 μm bending mode. The feature is unfortunately not very sensitive to changes in the mineralogy; therefore, it is useful to identify the presence of water, but less useful to constrain compositional information on the water-bearing mineral (Roush *et al.*, 1993). Other features in this spectral region, although weaker, are more diagnostic. The region between 2.1 and 2.5 μm shows a group of sharp but weak bands. These are caused because of the stretching fundamental vibrations of cation OH⁻, the 1.9 μm overtone of the H–O–H bending mode, and the 1.4 μm OH⁻ overtones in both H₂O and OH⁻ (Roush *et al.*, 1993; Salisbury *et al.*, 1991). The fine structure in this region can provide detailed information on the host minerals, because the combining spectral features are due

to different physical processes. The 1.9 μm overtone, for example, is caused by adsorbed or bound water in a mineral, like interlayer H₂O in a smectite clay. The presence of the 1.4 μm overtone, combined with the absence of the 1.9 μm absorption feature, indicates the presence of OH⁻-bearing minerals.

10.11.2.1.2 Gamma-ray and neutron spectroscopy

Gamma-ray and neutron spectroscopy can be used to survey planetary bodies for hydrogen. The abundance of hydrogen is an indirect measurement of the abundance of H₂O or OH⁻ in the surface layer of a planetary body under certain assumptions. While the detection of hydrogen by these methods is, in contrast to optical spectroscopy, unambiguous, the conversion to water abundance is strongly model dependent.

Galactic cosmic rays that penetrate the surface layer of a planetary body begin to lose energy immediately through a variety of mechanisms, mainly intranuclear cascades (Bertini, 1969). The secondary particles formed continue to lose energy by a variety of processes. The net result is that neutrons dominate the residual particle population below 100 MeV (Drake *et al.*, 1988; Feldman *et al.*, 1993) and gamma rays produced by the de-excitation of nuclei with characteristic energy spectra. On entering the surrounding material the generated neutrons are moderated by elastic and inelastic collisions. The resulting spectrum below 10 keV has two distinct energy regimes, the epithermal and the thermal neutrons.

The cross-section of hydrogen for the collision with neutrons is very high and therefore hydrogen is a very good neutron moderator. The epithermal neutron energy range (0.4–500 keV) is most sensitive for the detection and mapping of hydrogen (Feldman *et al.*, 2002). Simultaneous measurements of the flux of fast, epithermal and thermal neutrons emitted from a planetary surface allows constraining the amount of hydrogen as well as the vertical distribution in the subsurface. Deriving these constraints requires a multistep modeling approach, including the modeling of the neutron generation, the moderation in the subsurface, and propagation to orbit. (Feldman *et al.*, 1993). Finally a parametrization of the neutron energy distribution based on altitude of the observation and depth below the surface has to be derived (Feldman and Drake, 1986; Drake *et al.*, 1988). The model of the depth dependency includes assumptions about the soil composition, which add to the uncertainty of the obtained results.

The detection of hydrogen by gamma rays is based on the characteristic line at 2.2233 MeV generated by the neutron capture reaction (Reedy, 1978). Deriving the abundance and vertical distribution requires again a multistep modeling approach. This modeling again relies on assumptions about the soil composition and structure.

Gamma-ray and neutron spectroscopy allow detecting hydrogen buried in the subsurface. The penetration depths depend on the properties of the regolith and on the distance between detector and surface. Currently flying instruments such as the gamma ray spectrometer (GRS) suite on Mars Odyssey can detect hydrogen down to a burial depth of approximately 1 m (Feldman *et al.*, 2004). This is an advantage over visual and infrared (IR) spectroscopy which can detect water only at the surface. The disadvantage of the technique is a model-dependent ambiguity between abundance and burial depths.

10.11.2.1.3 Morphology

The solid surface of planets and moons is shaped by both endogenic and exogenic processes (Greeley, 1985). Therefore, its morphology is the result of their combined effects. In general, endogenetic processes concern tectonic and magmatic activities. The geologic action of water and ice belongs to the exogenic realm and can affect the surface morphology in several ways. Erosion, including transportation, tends to degrade the surface, while deposition of material leads to the formation of new sedimentary bodies. Erosional and depositional features result

from the nature and intensity of weathering and erosional processes, which, in turn, are related to climatic and environmental conditions.

Information about the morphology of planetary surfaces comes mainly from spacecraft images, but radar data (e.g., the Magellan data for Venus) and recently highly resolved topographic data from laser altimeters (e.g., Mars Orbiter laser altimeter (MOLA) data for Mars) also contributed much to our knowledge of extraterrestrial landforms. The study of landforms is called geomorphology. It is not the intent of this paper to give an overview of this discipline; instead, the reader is referred to classic textbooks, for example, Twidale (1976), Bloom (1991), Summerfield (1991), and Easterbrook (1999). Typically, processes related to water are classified depending on the environment: fluvial processes are related to running water, lacustrine and marine processes occur in lakes and seas, and glacial processes shape the surface by the action of ice. Water can also alter rocks and therefore change their physical properties, which in turn affects their susceptibility to other geological processes.

A major problem of the interpretation of remotely sensed data lies in the distinction whether erosion and sedimentation is driven by water or another geologic agent like wind. For example, the sinuous rilles on the Moon have been interpreted on the basis of Earth-based telescopic images and Lunar Orbiter photographs as erosional features carved by water by some early researchers (e.g., Urey, 1967). Zimbelman (2001) points out that the improvement in spatial resolution alone, as provided by Lunar Orbiter compared to Earth-based telescopic images, could not rule out several of the published hypotheses that are now known to be wrong. Only Apollo images in combination with *in situ* observations by astronauts together with laboratory investigations of returned samples revealed the volcanic origin of the rilles. Similarly, there was a long debate on Mars whether the huge outflow channels (see Section 10.11.3.3) were carved by water or another fluid like liquid CO₂, lava, or wind. It is therefore critical to test several hypotheses before coming to a final conclusion about the nature of morphological features.

The first and for a long time the strongest evidence that Mars has once experienced a warmer and wetter climate came from image data, starting with the Mariner 9 mission of 1971 (e.g., Masursky, 1973) and the Viking Orbiter missions, and their morphologic interpretation. Since water-related surface features often have a limited spatial extent and the

minimum object size to be detected in images depends on the available image resolution (see Zimbelman (2001) for a discussion on image resolution and its role in the photogeologic interpretation), the recent advance of high and very high resolution images taken by the MGS, Mars Odyssey, Mars Express, and Mars Reconnaissance Orbiter (*MRO*) missions has greatly enhanced our ability to analyse small-scale morphologies.

Except Mars and Earth, no other terrestrial planet displays morphologic features which can be attributed to water or ice. Sinuous rilles on the Moon were initially interpreted as water-related features, but it is now known that they were incised by flowing lava. Channels on Venus were probably formed in a similar way. Surface processes involving water imply an atmosphere, so the lack of it is not surprising on the Moon and Mercury. Venus has an atmosphere, but it contains no significant amounts of water vapor and it is too warm to allow liquid water or ice.

10.11.2.1.4 Radar

Radar for planetary exploration are used as ground-based radar, as ground-penetrating radar (GPR) on orbiters or rovers or as synthetic aperture radar (SAR) on orbiters.

Radar at frequencies commonly used for GPR in planetary exploration cannot unambiguously identify water. Dielectric-relaxation responses can uniquely identify water and ice, but these lie at frequencies of $>10\text{ GHz}$ and $<10\text{ kHz}$, respectively, well above and below the GPR radar band, and above 10 GHz the depth of investigation will be far less than 1 m.

At radar frequencies, rock and ice have representative dielectric constants of 5–10 and 3, respectively. In contrast liquid water near freezing has a dielectric constant of 87 (Fletcher, 1970; Hasted, 1973; Olhoeft, 1981; Heggy *et al.*, 2001). The dielectric constant of ice exhibits a stronger temperature dependence than water, whereas rocks show almost no temperature effect. The strong difference in the dielectric constants for rock and water makes radar a favorable technique for detecting water, especially if low-frequency sounding ($<10\text{ kHz}$) is used which gives access to the dielectric-relaxation response (Grimm, 2002). This is however only true if water is confined in a larger volume – for example, in an aquifer. If water is dispersed in the pore space of a dry regolith the contrast in the dielectric constant diminishes quickly. Assuming a soil–water mixture with a porosity of 5% yields a dielectric constant of 7–9 indistinguishable from dry rock (Shivola, 1996).

The small dielectric constant of ice allows a good penetration for radar signals. The difference to the typical values for rocks can allow detecting interfaces between pure ice and dry soil. Mixing of ice and water results in the same problem as described above for water–rock mixtures. Even large amounts of ice dispersed within the pore space are indistinguishable from dry rock and can therefore not be detected by GPR.

Ground-based radar observations of planetary surfaces assess the reflective properties of the surface and the near-surface material. Deposits of frozen volatiles on the surface, especially water ice, have a much lower transmission loss than silicate rocks. They produce therefore a higher average radar reflectivity than silicate rocks. This is however not a unique identification because the nature of the surface, especially the spatial-scale roughness of the radar wavelengths can significantly influence the reflective properties of the surface. The polarization properties of the surface can provide another indication for the existence of ice on or near the surface. Ice-containing surface tends to show a smaller degree of depolarization of the radar signal than dry rock surface. The higher reflectivity and the smaller depolarization effect are due to the details of the multiple scattering at a semitransparent surface (Hapke and Blewett, 1991).

10.11.3 Water on the Surface

If we focus on the surface and near-surface region of terrestrial planets, there are several possible reservoirs for water: in the regolith as ice or adsorbed, on the surface as ice or as standing or flowing body of water.

There are direct spectroscopic observations of water ice on Mars' polar caps as recorded recently by visible and infrared mineralogical mapping spectrometer (OMEGA; Observatoire pour la Minéralogie, l'Eau, les Glaces et l'Activité) and planetary Fourier spectrometer (PFS) on Mars Express (Bibring *et al.*, 2004b; Hansen *et al.*, 2005) and evidence for its existence today in permanently shadowed regions around the poles of Mercury and the Moon (Harmon and Slade, 1992; Butler *et al.*, 1993; Feldman *et al.*, 1998; Nozette *et al.*, 1996). So far no direct evidence for liquid water at present time has been found on any terrestrial planet. Mars, however, shows evidence that it had in its past large standing and flowing bodies of water. There are

also indirect indications for the existence of larger amounts of water in the very early history of Venus.

10.11.3.1 Valley Networks

Valley networks, first detected in Mariner 9 imagery (Masursky, 1973), are in contrast to the large outflow channels smaller in size and show a large variety of different morphologies. More than 1000 of individual valley networks were observed on Mars, and their similarity to terrestrial fluvial landforms implies that they were formed by water. They are the most striking morphologic evidence that water once flowed across the Martian surface. Nearly all valley networks are located in the heavily cratered highlands which led to the assumption that they were formed early in the Martian history. Although it is commonly accepted that water formed the valley networks, there is a long ongoing debate about the process (groundwater vs surface runoff), the source of the water (precipitation as rain or snow vs groundwater or ground ice), and therefore, if this points to a warm and wet, or cold and dry early Mars.

10.11.3.1.1 Morphology and formation

First it is important to note that the term valley network does not describe a river or stream channel. A valley is usually eroded by the action of flowing water. Interior river channels, the conduit in which a river flows, are rare on Mars. This might be due to erosion and sedimentation over long timescales. More recently, some interior channels were observed in high-resolution Mars Orbiter camera (MOC), thermal emission imaging system (THEMIS), and high-resolution stereo camera (HRSC) imagery (Malin and Carr, 1999; Malin and Edgett, 2000a; Irwin *et al.*, 2004, 2005b; Mangold *et al.*, 2004; Jaumann *et al.*, 2005).

Valley networks can reach lengths from <5 km to nearly 1000 km and widths from <1 km to nearly 10 km (Mars Channel Working Group, 1983). They can be broadly divided into large and small valley systems (Baker *et al.*, 1992): large valley systems, also termed as longitudinal valley systems (examples are Nirgal, Nanedi, Bahram, Ma'adim and Al Qahira Valles) are several hundred kilometres long and several kilometers wide. They exhibit upper reaches with short theatre-headed tributaries and broad flat-floored lower reaches with only a few tributaries. Small valley networks show a larger morphological variety. Most of them show a dendritic pattern in which tributaries converge downstream.

Many valley networks on Mars show morphologically an abrupt headward termination, theatre-like valley heads, prominent structural control, low junction angles, a relatively constant valley width downstream (Goldspiel *et al.*, 1993), and rectangular cross-sections with flat floors. More detailed descriptions about the valley network morphology are given in Baker (1982), Baker *et al.* (1992), Carr (1996), and Gulick (2001). The drainage densities are very low compared to terrestrial valleys (Carr and Chuang, 1997). Only some valley systems on volcanoes have comparable drainage densities to fluvial networks on Earth (Gulick and Baker, 1989). Comparative studies of valley networks in layered sedimentary rocks on Earth (Laity and Malin, 1985; Howard and Kochel, 1988) revealed the same morphology as well as low drainage densities. In contrast to Earth, where valleys are in most cases eroded by overland flow, the morphological attributes of Martian valleys and low drainage densities on Mars are best explained by groundwater sapping processes. Laity and Malin (1985) defined sapping process as

The process that leads to the undermining and collapse of valley head and side walls by weakening or removal of basal support as a result of enhanced weathering and erosion by concentrated fluid flow at a site of seepage.

However, the variety in morphology of Martian valley networks suggests that groundwater sapping was not the only process and surface runoff played an important role (Baker *et al.*, 1992; Craddock and Howard, 2002). Furthermore, even for sapping valleys the weathered and collapsed sediment have to be removed by surface runoff in order for retreat of the headwall (Craddock and Howard, 2002). As seen before from Viking imagery the inspection of valley networks with MOLA topographic data revealed that V-shaped cross-sections are common, typical of surface runoff (Williams and Phillips, 2001).

Recent observations in high-resolution imagery revealed dense, dendritic valley networks (Mangold *et al.*, 2004; Quantin *et al.*, 2005) which resemble terrestrial river networks formed by surface runoff processes (Figure 4). Furthermore, measurements based on higher-resolution imagery data and topographic data have increased the drainage densities drastically, which now reaches lower values of terrestrial river systems (Irwin and Howard, 2002; Hynek and Phillips, 2003; Mangold, 2004; Stepinski and Stepinski, 2005; Ansan and Mangold, 2006).

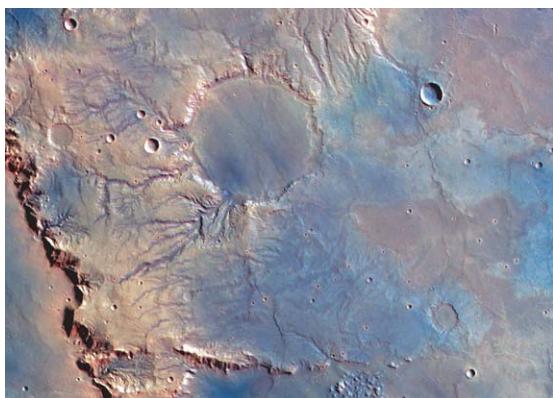


Figure 4 Dendritic valley networks on Mars. The large crater in the center of the image has a diameter of 45 km (False-colour HRSC image h00532_0000 at 60° E and 12.5° S).

10.11.3.1.2 Ages

Most of the valley networks are located in the oldest terrains of the southern highlands, which lead to the assumption that they were formed early in the Martian history (Pieri, 1976; Carr and Clow, 1981). Age determinations by crater counts are difficult due to the small areas of valley floors and subsequent obliteration of craters by infilling processes, such as eolian deposition and mass wasting. The age for some larger valley systems have been determined by crater counts on channel floors to be Noachian (Masursky *et al.*, 1977). The age of larger amounts of small valley networks were determined by superposition relationships on geologic units. Scott and Dohm (1992) analyzed the transection relationships of about 400 valley networks on geologic units. They derived maximum stratigraphic ages for most valleys between the Noachian and Early Hesperian. They suggest that possibly 30% might be Early Hesperian in age or even younger. Carr (1995) analyzed more than 800 valley networks with the same method and derived that ~92% have Noachian, ~4% Hesperian, and ~4% Amazonian maximum ages. However, these studies suggest that most valley networks were formed early in the Martian history, primarily in the Noachian and early Hesperian. Carr (1996) points out that the cratered highland valley networks are probably mostly Noachian in age and Hesperian-aged ones primarily occur where there is substantial local relief, such as in the Valles Marineris region and along the dichotomy boundary. Most Amazonian aged valleys are located on volcanoes (Carr, 1995; Scott and Dohm, 1992; Gulick and Baker, 1990). Recent studies support the view of an

extensive fluvial denudation in the Noachian with a declining rate into the Early Hesperian (Irwin *et al.*, 2005a; Howard *et al.*, 2005; Harrison and Grimm, 2005).

10.11.3.1.3 Source of water and climate

As pointed out by Craddock and Howard (2002), groundwater flow, recharge, and erosion are closely related to precipitation and surface runoff on Earth. Although sapping is an efficient process in loose sediments, there is no clear evidence from recent studies on Earth that groundwater processes cause significant erosion in resistant rock such as basalt (Lamb *et al.*, 2006). Due to bedrock strength, enormous discharges of water would have been necessary to erode the Martian valley networks (Howard and Kochel, 1988). Estimates of water volumes required for the formation of the valley networks indicate that the aquifers contained about 100 to 1000 times less water than needed to erode the valleys (Gulick, 2001). A recharge of the drainage areas was therefore necessary (Howard and Kochel, 1988; Goldspiel and Squyres, 2000; Grant, 2000; Gulick, 2001; Craddock and Howard, 2002). The simplest recharge mechanism is by precipitation and subsequent groundwater sapping and/or surface runoff, which would require higher atmospheric temperatures and pressures and point to a warmer and wetter climate on early Mars (e.g., Baker *et al.*, 1991; Carr, 1996; Craddock and Howard, 2002). Phillips *et al.* (2001) suggest a release of carbon dioxide and water from Tharsis magmas which may have sustained a warmer climate for the formation of valley networks around or after the end of the Noachian epoch.

Models of valley networks formation under dry and cold climatic conditions include the melting of dusty snowpacks by insolation (Clow, 1987), basal melting of snowpacks (Carr and Head, 2003), or melting of snowpacks by internal heat (Zent, 1999). Clifford (1987) proposed an aquifer recharge by basal melting of the South Polar Cap. However, this model is not consistent with many higher elevated valleys than the base of the South Polar Cap (Carr, 2002). Another suggested recharge model is the global hydrothermal circulation model of Clifford (1993), in which upward vapour flow driven by the geothermal heat gradient might form an icy permafrost layer. Subsequent melting of permafrost by geothermal heating might cause groundwater sapping (e.g., Squyres and Kasting, 1994).

The genesis of the younger valleys on volcanoes might have been different from valley networks in

the highlands. [Gulick \(2001\)](#) suggests that a hydrothermal system associated to magmatic intrusions replenishes itself. Also basal melting of a summit snowpack triggered by the heat flux of magmatic intrusions is proposed for the formation of fluvial valleys on Hecates Tholus ([Fassett and Head, 2006](#)).

10.11.3.2 Sedimentary Processes

10.11.3.2.1 The orbiter view

Paleolakes with their morphological characteristics like delta-like deposits, wave-cut terraces, shorelines, and evaporites support the hypothesis of past standing bodies of water on Mars.

Paleolacustrine features including deltaic sedimentary bodies and possible evaporitic deposits on Mars have been described in several studies using Viking imagery ([De Hon, 1992](#); [Forsythe and Zimbelman, 1995](#); [Wharton et al., 1995](#); [Newsom et al., 1996](#); [Cabrol et al., 1996](#); [Grin and Cabrol, 1997](#); [Cabrol et al., 1998](#); [Ori and Mosangini, 2000](#); [Ori et al., 2000a, 2000b](#)). About 200 paleolakes were identified by [Cabrol and Grin \(1999\)](#) on a global scale in impact craters on Mars, based on fluvial network inflows and morphologic lacustrine characteristics. Using stratigraphic relationships derived from geologic maps and crater-counting techniques, [Cabrol and Grin \(2001\)](#) estimated most of the paleolakes to be active in the Hesperian. However, a large number of lakes might have been active during the Amazonian, often associated with outflow channels ([Cabrol and Grin, 2001](#)).

Recent image and topographic data allowed detailed analyses of paleolakes, delta, and fan deposits, and allow the reconstruction of local hydrological cycles which helps to understand the Martian climatic history. Deltas exhibit a variety of morphologies ranging from Gilbert-type deltas which are formed when short-headed fluvial systems debouch in a lacustrine basin with a steep margin to shallow-water deltas by long fluvial systems showing a large number of distributary channels. However, there are also a number of fan-shaped bodies which often can be ascribed to alluvial fan due the lack of evidence of a lacustrine environment. [Irwin et al. \(2005a\)](#) found indications for deposition of pristine deltas and fans during a late stage of contributing valley entrenchment from the late Noachian to Early Hesperian. This is broadly supported by several local studies of paleolacustrine environments, although there are indications that lacustrine

environments existed within the Early Hesperian. A detailed study of the Tyras Vallis delta deposit in the Xanthe Terra region revealed a complex crater lake history with changing lake levels and water supplies between the Late Noachian and the Hesperian ([Di Achille et al., 2006a](#)). [Mangold and Ansan \(2006\)](#) also found evidence for a long-lasting Hesperian lake activity in the Thaumasia region from the analysis of a delta deposit and adjacent valleys. Deltaic, coastal, and lacustrine environments occurred during the Hesperian in the Holden crater, followed by glacial processes in Amazonian ([Pondrelli et al., 2005](#)). Sedimentary deposits in the Coprates Catena have been interpreted as an alluvial fan ([di Achille et al., 2006a](#)) or delta deposit ([Weitz et al., 2006](#)). Further evidence for Hesperian-aged lacustrine deposits with recent data has been found within Shalbatana Vallis ([Di Achille et al., in press](#)). A well-studied example of a possible Late Noachian ([Moore et al., 2003](#)) fluvial distributary fan occurs in the Eberswalde crater ([Figure 5](#)). It shows evidence for channelized and meandering flow indicating persistent flow conditions and formation of intracrater layered sedimentary sequences within fluvial and perhaps lacustrine environments ([Malin and Edgett, 2003](#)). The morphology of the layered deposit has been interpreted as a fan delta ([Bhattacharya et al., 2005](#); [Irwin et al., 2005a; Pondrelli et al., 2006](#)) or alluvial fan ([Malin and Edgett, 2003; Jerolmack et al., 2004](#)).

[Cabrol and Grin \(2001\)](#) found little evidence for a Noachian lacustrine activity. This contradicts the Noachian fluvial activity where lacustrine environments would be likely. They suggested that evidence for it might have been lost due to erosion and major resurfacing episodes later in Martian

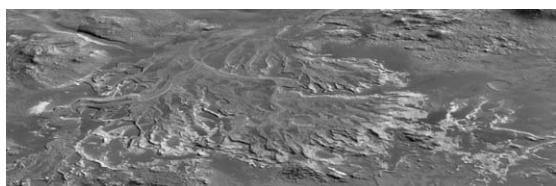


Figure 5 The Eberswalde crater hosts the best-known example of a deltaic structure on Mars. Evidence for channelized and meandering flow indicates persistent flow conditions and formation of intracrater layered sedimentary sequences within fluvial and perhaps lacustrine environments ([Malin and Edgett, 2003](#)). The image is a perspective view from the east, the image width is about 20 km (MOC image mosaic draped over topography derived from HRSC stereo images).

history. Possible remnants of this period might be archived in numerous layered sedimentary outcrops which have been observed in MOC imagery (Malin and Edgett, 2000b). These features consist in most places of (1) thin, horizontally bedded layers at the base, (2) a thick massive unit, which is not layered or poorly bedded in between, and (3) a thin mesa unit at the top (Malin and Edgett, 2000b). It is unclear if they have been formed under subaerial or subaqueous conditions and the source regions and transport paths are not preserved, but the stratigraphic positions as well as exhumed craters suggest that the outcrops have been formed most probably in the Noachian (Malin and Edgett, 2000b).

10.11.3.2.2 The view of the MER rovers

Aqueous processes have also been inferred from morphological observations by the MER, Opportunity, in Meridiani Planum (Squyres *et al.*, 2004). The layered rocks consist of sandstones, which show several sedimentary structures like cross-bedding, festoon cross-lamination, resistant fracture fills, and polygonal forms (Grotzinger *et al.*, 2005). Although the overall environment was probably dominated by eolian processes, these structures, together with geochemical and mineralogical evidence (see Section 10.11.5.1), indicate that abundant acidic groundwater was present, that acidic and oxidizing conditions prevailed, and that inundations led to episodic shallow subaqueous flows of liquid water (Squyres *et al.*, 2006).

Several terrestrial analogs, ancient and modern, have been proposed to illustrate the paleoenvironment in Meridiani Planum. A summary is given by Benison and Bowen (2006), who favor ephemeral acid saline lakes in SW Australia as the best-known Earth analog. There, shallow lakes and adjacent groundwaters have low pH values (1.5–4) and minerals are precipitated that are also found in Meridiani Planum (gypsum, halite, iron oxides, jarosite). Flooding, evaporation, desiccation, and winds dominate the environment (Benison and Bowen, 2006).

10.11.3.3 Outflow Channels

Outflow channels, first detected in images of the Mariner 9 spacecraft (Masursky *et al.*, 1973), are the most spectacular examples of Martian surface features formed by the erosive power of water. Their dimensions are much larger than those of other fluvial landforms on Mars as, for example, valley networks (Section 10.11.3.1). The largest outflow channels can reach lengths of thousands of

kilometres, widths of tens to hundreds of kilometres, and depths of more than 1 km. They are not uniformly spread over the Martian surface, but cluster in a few distinct regions, in particular around the Chryse Planitia basin in the eastern vicinity of the huge magmatotectonic center of Tharsis (Figure 6(b): circum-Chryse outflow channels). Other concentrations occur in Elysium Planitia, in the eastern Hellas region, and on the southern margin of Amazonis Planitia. Outflow channels formed episodically in several stages over a significant part of the geologic history of Mars, beginning in the Late Noachian. The main activity seems to have been in the Hesperian period, with lesser stages continuing well into the Amazonian period (see table IV of Clifford and Parker, 2001; Neukum and Hiller, 1981; Scott and Tanaka, 1986; Greeley and Guest, 1987; Lanz, 2004; Werner, 2006). Two outflow channels in the Elysium region, Marte Vallis, and the Athabasca Valles, even show evidence for some of the youngest fluvial activity on Mars (Werner *et al.*, 2003; Burr *et al.*, 2002a, 2002b; Burr, 2003).

Outflow channels are important in at least two aspects for the Martian history: first, their enormous sizes strongly indicate that large volumes of liquid water or water ice once flowed across the surface of Mars. Secondly, outflow channels would have been the most obvious pathways for water that filled the northern lowlands to form an ancient Martian ocean (Baker *et al.*, 1991; see also Section 10.11.3.4).

10.11.3.3.1 Morphology

Outflow channels start full size at discrete sources. These sources may be fractures (e.g., at Athabasca or Mangala Valles; Figure 7: Mangala), but most commonly they are located in the so-called chaotic terrains (Figure 6(d): Aromatum Chaos/chaotic terrains). These zones are characterized by large blocks of plateau material and intermediate depressions, often arranged in a ‘chaotic’ geometry. They seem to have formed by the collapse of plateau material that, where undisturbed, currently forms the surroundings of the chaotic terrains. The most characteristic landforms associated with outflow channels themselves are streamlined or ‘teardrop-shaped’ islands, which were clearly carved by a fluid medium and indicate the flow direction (Figure 7: Mangala). The overall pattern of outflow channels is often braided or anastomosing (Figures 6(b) and 6(c): Outflow channels in MOLA). Other surface features typically observed at the floors of outflow channels are bed forms like scour marks,

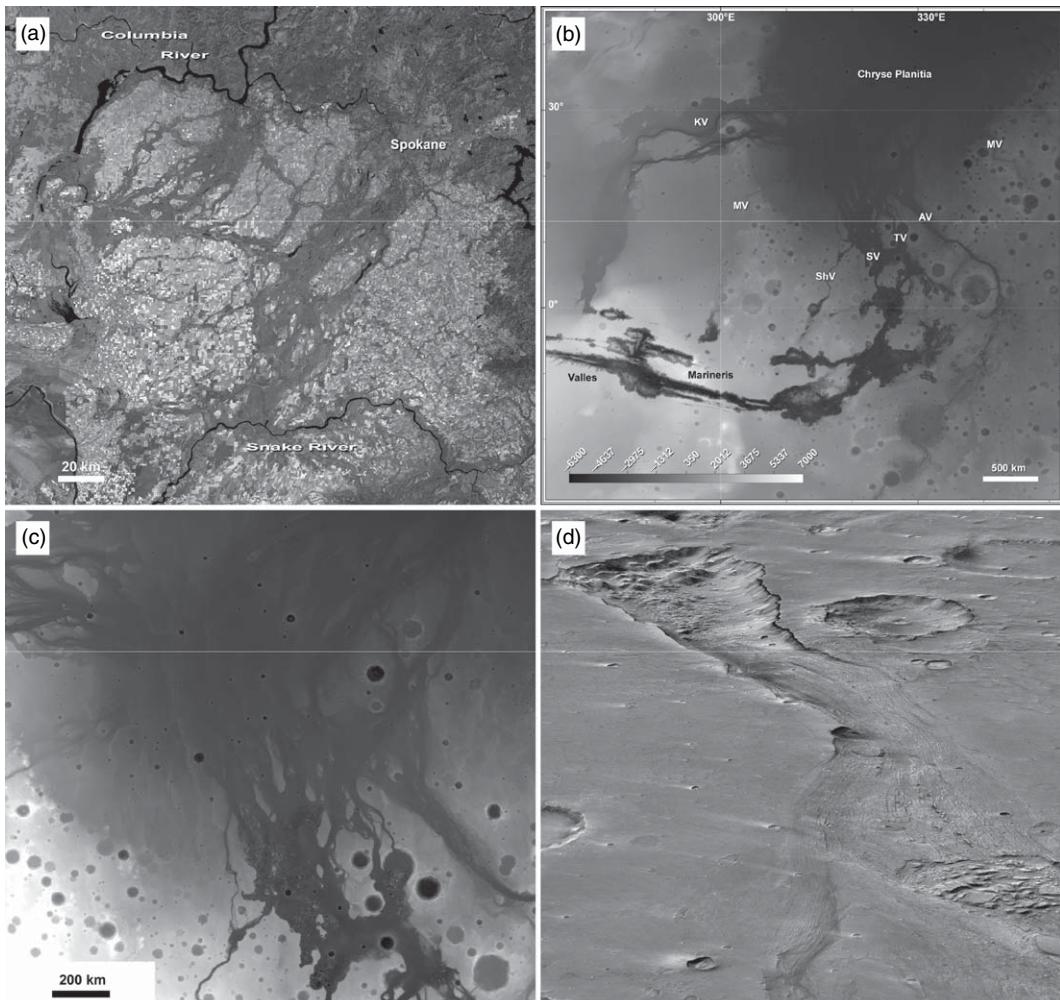


Figure 6 Outflow channels are the most dramatic geomorphic evidence for the flow of liquid water or ice on the Martian surface. They were interpreted as the erosional products of catastrophic floods on the basis of the comparison to terrestrial analogs. (a) The channeled scablands in Washington State (USA) formed when an ice-dammed lake drained in a huge flood toward the Pacific Ocean (Bretz, 1923). The channel floors can be recognized as dark areas and form an anastomosing pattern (Landsat mosaic; North is up). (b) Outflow channel on Mars. The largest channels debouch into the topographic low of the Chryse Planitia basin (KV: Kasei Vallis; MV: Maja Vallis; ShV: Shalbatana Vallis; SV: Simud Vallis; TV: Tiu Vallis; AV: Ares Vallis). The image is a gridded Digital Elevation Model derived from MOLA altimetry data (see scale bar for elevations). (c) Detail from (b), showing the anastomosing nature of the downstream parts of some channels. Note the similarity to the terrestrial analog shown in (a), but also note the difference in scale. (d) Perspective view of an outflow channel. The source in the left background is a typical chaotic terrain, Aromatum Chaos, and the outflow channel, Ravi Vallis, extends from there to the right foreground, displaying a grooved channel floor (HRSC image mosaic draped over MOLA topography; the diameter of the large crater is about 26 km).

grooves, cataracts, and possible sinkholes ('kolks') (Baker, 1982), testifying that the outflow channels are true river channels (as opposed to valleys; see discussion in Carr, 1996). Outflow channels have only few tributaries and maintain a more or less constant width downstream. More detailed descriptions of the channel morphologies are given by Baker (1982), Baker and Nummedal (1978), and Baker *et al.*

(1992). It was recognized early in the interpretation of spacecraft images that the collective association of these landforms bears a striking similarity to the so-called Channeled Scablands in the northwestern USA (Baker and Milton, 1974; Baker, 1979, 1982; Baker and Kochel, 1979) (Figure 6(a): Landsat). These giant erosion features were interpreted in a series of pioneering studies by Harlen Bretz early in

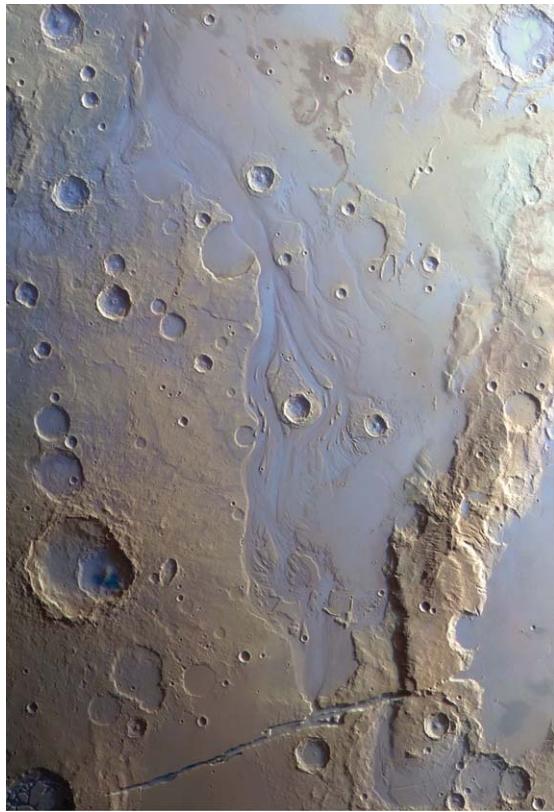


Figure 7 Mangala Vallis, a typical outflow channel on Mars. The channel originates at a tectonic fracture at the bottom center of the image and flowed toward north (top of image), eroding streamlined islands. False-color HRSC image 0644 (image width is about 370 km).

the twentieth century as results of the catastrophic drainage of ice-dammed lakes during the latest glaciation in the Pleistocene (e.g., Bretz, 1923). Their significance extends well beyond the Channelled Scablands in that they helped to understand that catastrophic events must not be neglected in the deciphering of the Earth's geologic record ('Uniformitarianism' vs 'Catastrophism'; for a comprehensive modern view on this debate see Baker and Nummedal (1978)). Similar 'spillways' have later been interpreted by means of modern spacecraft data in Siberia (Manych spillway, Turgay spillway; e.g., Baker *et al.*, 1993; Grosswald, 1998).

The cataclysmic flood hypothesis is, however, not without difficulties, as pointed out by Baker *et al.* (1992). First, under current climatic conditions water is not stable. Since the floods were probably very short-lived, this does not seem to be a major problem. Second, there are no obvious fluvial deposits like deltas at the termini of the channels. The third

problem is that the source areas, mostly the chaotic terrains, are relatively small when compared to the amount of water which probably carved the outflow channels. This discrepancy might be explained by the headward migration of the source areas as more and more terrains collapse in the subsequent stages of flooding (Baker, 1982).

10.11.3.3.2 Morphometry/discharge

Many attempts have been made to estimate the peak drainage in the outflow channels from the geometry of the paleochannels. The first and for a long time the most accepted approach was the application of modified versions of the empirically determined Manning equation (Manning, 1891), which predicts flow velocities in terrestrial channels. In two outstanding papers Komar (1979, 1980) pioneered studies of this kind and found values in the range of 10^7 to $10^8 \text{ m}^3 \text{ s}^{-1}$ (for Maja, Ares, and Tiu Valles, assuming a flow depth of 100 m). Even higher values of up to $10^9 \text{ m}^3 \text{ s}^{-1}$ have been reported for Kasei Vallis, the largest outflow channel on Mars (Robinson and Tanaka, 1990; Baker, 2001). Williams *et al.* (2000) used improved topographic data from the MOLA instrument and reported values between $8 \times 10^4 \text{ m}^3 \text{ s}^{-1}$ and $2 \times 10^7 \text{ m}^3 \text{ s}^{-1}$ for channels of the Kasei Vallis system. It has to be emphasized, however, that Komar (1979, 1980) was aware of the shortcomings of this approach (see discussions in Wilson *et al.* (2004) and Kleinhans (2005)). More recently, Wilson *et al.* (2004) derived a new value for the Manning n coefficient and found that previous analyses based on modified versions of the original Manning equation overestimated the water-flow velocities by a factor of up to 2.5. They recommend that their improved value for the Manning n coefficient should be used or, preferably, that a modern form of the Manning equation or the Darcy–Weisbach equation be used for future studies. In an up-to-date review of water flow and sediment transport models, Kleinhans (2005) also recommends the use of the Darcy–Weisbach equation and gives ranges of flow discharges for large outflow channels like Ares or Kasei Vallis in the order of 10^2 – $10^3 \times 10^6 \text{ m}^3 \text{ s}^{-1}$. Discharges within the same order of magnitude were found by Burr *et al.* (2002a) for the Athabasca Valles (1 – $2 \times 10^6 \text{ m}^3 \text{ s}^{-1}$), Ghatan *et al.* (2005) for Mangala Valles ($\sim 5 \times 10^6 \text{ m}^3 \text{ s}^{-1}$), Coleman (2004) for Ravi Vallis (1 – $35 \times 10^6 \text{ m}^3 \text{ s}^{-1}$), and Leask *et al.* (2006a), also for Ravi Vallis (between <1 and $30 \times 10^6 \text{ m}^3 \text{ s}^{-1}$). Leask *et al.* (2006a) calculated the duration of the flooding event from the volume of

the eroded material, from plausible assumptions on the sediment-carrying capability of water, the minimum total volume of water involved, and the crustal erosion rate. They obtained a duration of 2–10 weeks, comparable with a duration of 1–3 months for the flooding in the Mangala Valles, as reported by Ghatal *et al.* (2005).

10.11.3.3.3 Formation

Many fluid media have been proposed to explain the formation of outflow channels, for example, wind (Blasius and Cutts, 1979), lava (e.g., Carr, 1974; Leverington, 2004), liquid hydrocarbons (Yung and Pinto, 1978), liquid CO₂ (Milton, 1974; Hoffman, 2000a; Tanaka *et al.*, 2001), the dissociation of methane hydrate (Max and Clifford, 2001), or liquefaction and/or debris flows (Nummedal, 1978; Nummedal and Prior, 1981; Tanaka, 1997, 1999; Rodriguez *et al.*, 2006) (for a more complete list see Carr (1996) and references therein). However, the consensus between most researchers is that only liquid water or water ice are physically plausible candidates in the Martian environment (e.g., Carr, 1996; Baker, 2001; Coleman, 2003, 2005). While the collective assemblage of outflow channel landforms suggests, by analogy with the terrestrial Channeled Scablands, cataclysmic release of liquid water as the formation mechanism, it has to be emphasized that water ice can also explain some of the properties of outflow channels (Lucchitta, 1982, 2001).

How could huge volumes of water be released in short time on the Martian surface? Most researchers think that a pressurized aquifer in depth, covered by an impermeable crustal zone of ground ice (the cryosphere; Clifford (1993)), was tapped, and the artesian release of that pressurized water led to the catastrophic drainage of the aquifer, collapse of the surface to form chaotic terrains, and outflow channel formation (e.g., Masursky *et al.*, 1986; Baker and Kochel, 1979; Carr, 1979; Theilig and Greeley, 1979; Baker *et al.*, 1991). Some chaotic terrains are completely enclosed and do not have a surface connection to an outflow channel. On the other hand, there are zones of surface subsidence between some chaotic terrains and outflow channels (e.g., between Ganges Chasma and Shalbatana Vallis) (e.g., Cabrol *et al.*, 1997). Both observations indicate that channeled subsurface flow of water is an important factor in the mobilization of the water masses responsible for outflow channel generation (Carr, 1995). Rodriguez *et al.* (2003) suggest that extensive cavernous systems, triggered by intrusive magmatism into

permafrost, served as pathways for the migration of water that finally emerged into the outflow channels.

Several processes have been invoked to break the thick permafrost zone or cryosphere above a pressurized aquifer, including magmatic intrusions (e.g., Masursky *et al.*, 1977, 1986; Greeley and Spudis, 1981; Mars Channel Working Group, 1983; Mouginis-Mark *et al.*, 1984; Rodriguez *et al.*, 2003), and in particular dikes (e.g., Head *et al.*, 2003b) and sills (Leask *et al.*, 2006b), and tectonism (e.g., Cabrol *et al.*, 1997). Manga (2004) showed for the case of the recent flooding at Cerberus Fossae that large subsurface aquifers with permeabilities comparable to those of young basaltic aquifers on Earth can discharge enough water to account for the estimated 10⁶ m³ s⁻¹ (Burr *et al.*, 2002a), without the need for unusual physics or material properties. An alternative way to release water in large amounts might have been the catastrophic drainage of ice-covered lakes in the Valles Marineris area or in the chaotic terrains (e.g., McCauley, 1978; Robinson and Tanaka, 1990). Outflow channels often show evidence for multiple stages of flooding, which requires recharging of the aquifer(s). This could have taken place by basal melting of the topographically high-lying South Polar Cap (Clifford and Parker, 2001), supplying a globally connected aquifer with water, or by precipitation on the elevated Tharsis Bulge in times of higher obliquity (Harrison and Grimm, 2004).

10.11.3.4 Oceans

The Earth is topographically divided in two main regions on a global scale, the high continents and the low-lying, water-covered ocean floors. Mars is also topographically divided in two main regions on a global scale, the southern highlands and the northern lowlands. The lowlands north of the dichotomy boundary would represent the largest drainage basin on Mars (Banerdt and Vidal, 2001; Smith *et al.*, 2001), and most of the largest erosional features on Mars, the outflow channels (see Section 10.11.3), debouch into them. Not surprisingly, perhaps, a variety of studies investigated the possibility of an ancient large body of standing water in the lowlands.

10.11.3.4.1 Models involving oceans

Oceans as water reservoirs from which the atmosphere could be replenished constitute major elements in some models of the evolution of the Martian hydrology. Perhaps one of the most influential was a conceptual scheme outlined by Baker *et al.*

(1991). Overall, this model suggests that the climate since the end of the Noachian was mainly cold and dry, in agreement with the very low erosion rates since then (e.g., Golombek and Bridges, 2000; Golombek *et al.*, 2006). However, short episodes of enhanced hydrologic activity were triggered by volcanism in Tharsis and subsequent cataclysmic outbursts of water that was transported down the outflow channels and finally ponded in the lowlands as oceans, called ‘Oceanus Borealis’. The release of the greenhouse gases, water, and CO₂ would have created a short-lived denser atmosphere and an active and dynamic hydrological cycle, which involves precipitation and channel formation (e.g., on areas of high elevations like Alba Patera). In colder regions, glacial landforms would result from snowfall and glaciation. When the released water finally was trapped as ground ice or water, the hydrologic activity would have stopped. The short phases of enhanced hydrologic activities would also have been responsible for the formation of a variety of geomorphic features on Mars, which were otherwise hard to explain. Among these were possible glacial features in the Southern Hemisphere (Kargel and Strom, 1992) and young, Amazonian-aged valley networks on some volcanoes (Gulick and Baker, 1989, 1990). It has to be noted, however, that recently other possibilities to provide the water for the formation of these young valley networks have been investigated, like the melting of snow (e.g., Fassett and Head, 2006) that was deposited during periods of higher obliquity and enhanced precipitation in lower latitudes.

Most studies proposing an ocean explained it as the consequence of water release through the giant outflow channels. Therefore, its age was commonly assumed to be roughly coeval with outflow channel activity, that is, Hesperian (~3.7–3.3 Ga) or even Early to Middle Amazonian (~3.3–0.6 Ga) in age. In contrast, a recent theory of the hydrologic evolution of Mars postulates the existence of an ocean in the first billion years of the planet’s history: Clifford and Parker (2001) investigated the hydraulic conditions that were necessary to explain the outflow channel formation in the Late Hesperian by the catastrophic release of groundwater from confined aquifers. They find that a several-kilometer-thick cryosphere would have been required to support an elevated groundwater table at or above the source region of the outflow channels. In the Noachian, when the climate might have been warmer and the global heat flux was several times higher, the cryosphere would have been

too thin to prevent the water from accumulating in the regions of the lowest geopotential. The inevitable consequence would have been the formation of a primordial ocean in the Noachian, which might have covered a third of the planet’s surface (Clifford and Parker, 2001).

10.11.3.4.2 Assessment of geomorphic evidence

One of the first studies that speculated about the existence of an ocean was by Jöns (1984, 1985). He described possible sediments and mudflows (Jöns, 1984), and concluded that the northern lowlands might have hosted an ancient mud ocean fed by mass-wasting processes (Jöns, 1985, 1986). At about the same time, McGill (1985) and Lucchitta *et al.* (1986) mapped polygonal patterns in the northern lowlands and found a spatial association with the terminations of the big outflow channels. Lucchitta *et al.* (1986) suggested that the polygons developed in sedimentary layers and concluded that these might have been deposited by outflow channel discharge in an ancient ocean, which could have been covered by ice. Later, but still on the basis of Viking Orbiter images with a resolution of typically 60–100 m per pixel, Parker *et al.* (1989, 1993) made an attempt to map shorelines around the northern lowlands. They suggested two possible sets of shorelines, an outer set called ‘contact 1’ and an inner set (‘contact 2’). Based on their mapping and a discussion on possibly involved water volumes, Parker *et al.* (1993) propose that the northern lowlands have been occupied by an ocean fed by outflow channel activity, the shorelines representing highstands of the water level. For a long time, the shoreline maps by Parker *et al.* (1989, 1993) were the foundation of most studies speculating about northern Martian oceans. Later, Edgett and Parker (1997), Clifford and Parker (2001), and Webb (2004) reported on additional shorelines, some of them requiring significant volumes of water to fill the outlined depressions.

In these studies, the outflow channels would have been the pathways for the water filling the hypothesized ancient oceans. Therefore, the plausibility of the ocean hypothesis depends critically on the relationship between the outflow channels and the northern lowlands as the natural sink for their water.

The morphologic expression of six major outflow channels disappears in the downstream direction at about the same elevation around Chryse Planitia. Ivanov and Head (2001) note that this fact might suggest a common base level and would be in

agreement with large lakes or oceans in the lowlands in the Hesperian or in the Early Amazonian, the main periods of outflow channel activity (see Section 10.11.3.3). The Vastitas Borealis Formation (VBF), which covers lower Hesperian ridged plains in the lowlands, is interpreted to be an eroded residue of ice-rich sedimentary deposits, which were emplaced by the floods carving the outflow channels (e.g., Head *et al.*, 2002; Kreslavsky and Head, 2002; Tanaka *et al.*, 2001). The very similar ages of the outflow channels and the VBF supports this idea. Further support comes from the approximately identical volumes of the material eroded by the outflow channels and the VBF (Carr and Head, 2003).

Polygonal terrain in Utopia, previously suggested to be of sedimentary origin (Lucchitta *et al.*, 1986), was interpreted to be a tectonic consequence of flexural rebound of the crust after removing the load of a former ocean (Hiesinger and Head, 2000). Boyce *et al.* (2005) examined the morphometry of impact craters in the northern lowlands with MOLA data. A population of Hesperian-aged craters with a low depth-to-diameter relationship seems to be filled by material of the VBF. Boyce *et al.* (2005) conclude that the crater filling is the result of burial from an ice-rich mantle. They use the extent of this crater population as an indicator of the areal extent of the VBF and the possible limits of an ancient ocean in the Late Hesperian or Early Amazonian.

10.11.3.4.3 Tests of ocean hypotheses

More recent data from the MGS mission (Albee *et al.*, 2001) allowed to test the ocean hypothesis. MOLA data showed that the northern lowlands were extremely smooth and level at scales of a few hundred meters to a few kilometres (Smith *et al.*, 1998; Aharonson *et al.*, 1998; Kreslavsky and Head, 1999). Indeed, they can be compared to the smoothest terrestrial surfaces like the abyssal plains (Smith *et al.*, 1998). Head *et al.* (1998, 1999) used MOLA data to compare the smoothness of the surface within and outside of the mapped shorelines of Parker *et al.* (1989). The surface is rougher at all scales from several hundred meters to tens of kilometers outside of contact 1 than between contacts 1 and 2, and similarly the surface within contact 2 is smoother at all scales than outside contact 2. They also sampled the elevations along the shorelines, and found that elevations of contact 1 span almost 11 km, a value that is clearly inconsistent with any plausible equipotential surface in the past. Contact 2 still has an elevation range of \sim 4.7 km, but the major deviations

from an equipotential surface occur in Arabia and Elysium (>1 km downward deflection of the mapped features from an equipotential surface), and in Tharsis (>2 km upward deflection), both being regions where vertical tectonic movements might have occurred after the formation of the shorelines. Head *et al.* (1999) concluded that MOLA data are consistent with hypotheses of oceans in the Martian past. However, several objections can be raised. First, it seems questionable whether post-shoreline downward flexure in Arabia or uplift in Tharsis could really have reached these values in the order of kilometers. Although Clifford and Parker (2001) admit that errors in the original, Viking-based identification of shorelines by Parker *et al.* (1989) might have contributed to the wide range of elevations, significant tectonic forces would have been required after outflow channel activity and shoreline formation in the Late Hesperian or even later. Second, detailed analyses of the suggested shoreline features have not shown morphological evidence in very high-resolution MOC images. Malin and Edgett (1999) investigated MOC images crossing the proposed shorelines at several locations and found no evidence for coastal landforms. This result was challenged by Clifford and Parker (2001) who insist that some of the features shown by Malin and Edgett (1999) against shorelines are indeed evidence in favor of shorelines. However, Carr and Head (2003) in their assessment of the observational evidence for oceans also could not confirm the existence of coastal landforms. While most studies tried to find geomorphic evidence in the form of wave-cut erosional landforms or interior sedimentary features, Ghatan and Zimbelman (2005) searched the shorelines mapped by Parker *et al.* (1993) and Webb (2004) for coastal constructional landforms, but found almost none. Based on this paucity of geomorphic evidence for shorelines, Ghatan and Zimbelman (2005) do not exclude the possibility of a northern ocean, but put constraints on its nature. For example, the water flowing down the outflow channels could have ponded in smaller lakes and not in a global-scale ocean. Such ponds and lakes had already been mapped by Scott *et al.* (1995) on the basis of Viking Orbiter images, but a test of their existence, for example, by examining their potential shorelines using higher-resolution post-Viking data, has not been performed yet. It has to be noted, however, that the detection of shorelines in satellite imagery is difficult even on Earth, and that erosion over possibly billions of years would have obscured any such features. Moreover, an ocean would likely

freeze very rapidly (Carr, 1983; Baker, 2001; Kreslavsky and Head, 2002) before any significant coastal erosion by wind-generated waves could start.

If there were major bodies of standing water on ancient Mars, it would be expected that sedimentation of carbon dioxide derived from the atmosphere would have led to carbonate formation (e.g., Catling *et al.*, 1999). The lack of mineralogical evidence for carbonates has always been a major argument against long-lived oceans and lakes on Mars. The thermal emission system (TES) on board of Global Surveyor did not show spectral signatures of carbonates (Christensen *et al.*, 2001a; Stockstill *et al.*, 2005). Only small amounts are distributed in the Martian dust (Bandfield *et al.*, 2003). The OMEGA imaging spectrometer confirmed the results, reporting no carbonates above the instrument's detection limit (Bibring *et al.*, 2005). Moreover, the identification of olivine, a mineral which is easily destroyed by chemical weathering, might argue against long-lived reservoirs of liquid water at the surface of Mars. It has to be noted, however, that the lack of carbonates does not necessarily exclude the possibility of oceans on ancient Mars. Even minor amounts of sulfur dioxide would produce acidic waters and prevent the formation of carbonates, instead leading to sulfate deposition (Lane *et al.*, 2004; Fairén *et al.*, 2004). Calcium sulfates have indeed been found on Mars recently by Gendrin *et al.* (2005) and Langevin *et al.* (2005a).

While the ocean hypothesis is attractive to explain several observations related to the hydrologic evolution of Mars, the existence of an ocean is yet to be proved. Meanwhile, many key questions are still open (summarized by Ghatal and Zimbelman (2006)): was there a Noachian ocean as proposed by Clifford and Parker (2001)? was there a genetic relationship between outflow channel discharge and water ponding as an ocean in the northern lowlands during the Hesperian? was ocean formation a repeated phenomenon (Baker *et al.*, 1991)? was an ocean liquid, or did it freeze rapidly, and then sublimate? It seems questionable if new remote sensing data can answer these questions. So far, the Mars advanced radar for subsurface ionosphere sounding (MARSIS) radar on Mars Express did not reveal the existence of massive ice deposits beneath the surface of the northern plains, a result that, if confirmed, argues against the evolutionary model outlined by Clifford and Parker (2001). Given the old age of the hypothesized oceans and the susceptibility of coastal landforms to erosion, the search for coastlines with

imaging and topographic data is considered to be difficult. Even if the images of the high-resolution imaging science experiment (HiRISE) camera with a lateral resolution of 30 cm per pixel (McEwen *et al.*, in press) do not show any evidence, this result would not disprove an ocean. Perhaps the best chances to reach more definitive results will involve lander experiments (e.g., with GPR experiments), and sample return missions.

10.11.3.5 Glacial, Permafrost, and Ground-Ice Features

10.11.3.5.1 Glacial landforms

Glacial processes on Earth produce a wide variety of landforms at different scales (e.g., Embleton and King, 1975; Sugden and John, 1976; Hambrey, 1994; Bennett and Glasser, 1996; Benn and Evans, 1998; Evans, 2003; Menzies, 1995). Examples of glacial erosional features include striated rock surfaces at small scales, roches moutonées at intermediate scales, and cirques, troughs, and fjords at large scales. Well-known glacial depositional features are, for example, moraines and till deposits. As part of a glacial system, not only the processes at the glacier surface boundary are of importance, but also those landforms produced by fluvial discharge produced by, for example, melting of glaciers.

The surface of Mars has probably been a cold environment over most of the planet's history (e.g., Carr, 1996; Solomon *et al.*, 2005; Bibring *et al.*, 2006). This and the fact that water, although in unknown quantity, is assumed to have been present at the surface throughout most of this time raises the possibility that glaciers have formed on the Martian surface, and that glacial surface features are still observable today.

Mariner and Viking images showed landforms that were indeed interpreted as glacial in origin, but the number of studies was limited. Williams (1978) and Lucchitta (1981) described landforms resembling recessional moraines at the western and northwestern flanks of the large Tharsis shield volcanoes, which are also shown as possible glacial deposits on the global-scale map by Scott and Tanaka (1986). The morphology of some outflow channels is also indicative of glacial action (Lucchitta, 1982, 2001) (see Section 10.11.3.5), and both flood and glacial processes may have simultaneously or sequentially shaped them (Costard and Baker, 2001).

Intriguing evidence for glacial processes reported in the Viking-era are esker-like and moraine-like

landforms in both hemispheres (Kargel and Strom, 1992). Kargel *et al.* (1995) reported an assemblage of features in the northern plains that resemble terrestrial glacial landforms, including moraines, tunnel valleys, and possible eskers.

These landforms were recognized to be relatively young (Kargel and Strom, 1992). Since glaciers need to be recharged by precipitation and a water cycle, their formation late in Mars' history would require a major climate change on Mars relatively recently. The scenario of Baker *et al.* (1981) accounted for this problem in postulating the episodic formation of oceans by outflow channel activity (see Section 10.11.3.3), which then triggered short-lived episodes of a warmer climate with precipitation. It has to be noted, however, that the interpretation of these features as being glacial was partly met with skepticism (Carr, 1996), since the low amount of erosion in the Amazonian (e.g., Golombek and Bridges, 2000; Golombek *et al.*, 2006) seems to contradict major climate excursions late in Martian history.

Images with higher spatial resolution (MOC, THEMIS-VIS, HRSC) revealed more details of the moraine-like features. Head and Marchant (2003), Shean *et al.* (2005), and Milkovich *et al.* (2006) investigated the structures on the northwestern flanks of the Tharsis Montes (Arsia, Pavonis, and Ascraeus Montes) and Olympus Mons (Figure 8) and found an assemblage of landforms that they interpreted as the result of cold-based glaciation, in analogy with cold-based glaciers in Antarctica, which are frozen to the ground and do not move by sliding on the

ground, but by internal deformation. Bed erosion at cold-based glaciers is limited due to the fact that ice usually does not reach its melting temperature. Hauber *et al.* (2005) detected possible glacial features at the northwestern flank of Hecates Tholus, where previously no glacial features had been identified. Additional evidence for mid-latitude glaciation was described by Head *et al.* (2005, 2006).

A particular type of ice-related morphology exists near the Martian equator in Elysium Planitia. Rafts of darker material seem to have flown on a brighter matrix. In places, the dark material obviously broke into smaller pieces. On the basis of its overall similarity to terrestrial shelf-ice, Brakenridge (1993) suggested that quite recent surface water releases filled a topographic depression, that ice floes formed almost immediately, and that either grounded ice or an ice-covered sea still persists. Later, the area was revisited by Murray *et al.* (2005), who analyzed HRSC images and basically came to the same conclusions. The massive water release required to fill the roughly North Sea-sized basin would imply a large underground reservoir. It seems possible that water was transported through the Athabasca Valles, a young outflow channel, towards Elysium Planitia. It must be noted, however, that the pack-ice theory for the platy landforms in Elysium Planitia is under debate. An origin by flowing lava, which cools at the surface and builds a solidified upper crust that can break in a way similar to pack-ice also seems to be possible (Keszthelyi *et al.*, 2004).

10.11.3.5.2 Permafrost and ground-ice landforms

Permafrost on Earth is defined as soil or rock that remains below 0°C for two or more consecutive years, regardless of the water content. Following this definition, Mars may well be regarded as a permafrost planet. Beneath a thin surficial layer, which is affected by the diurnal temperature wave, all of its upper crust can be considered to have temperatures below 0°C. Ground ice on Earth is ice in frozen ground. Both permafrost and ground ice belong to the realm of the periglacial zone. The term periglacial was first introduced by Lozinski (1909), and in 1910 the concept of a 'periglacial zone' was established to describe climatic and geomorphic phenomena occurring peripheral to the Pleistocene ice sheets (e.g., Huggett, 2003). The term 'periglacial' is nowadays used in a wider sense and periglacial areas are not restricted anymore to any location or time; it is generally qualitatively used

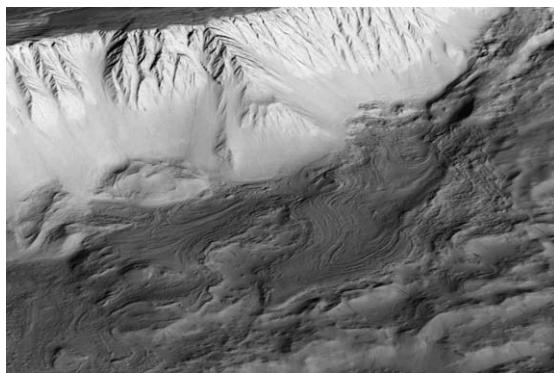


Figure 8 Lobate deposits at the foot of the western scarp of Olympus Mons. Such deposits were interpreted as possible glacial landforms by Lucchitta (1981). Milkovich *et al.* (2006) ascribe their origin to episodes of increased obliquity during the Late Amazonian period of Mars. Perspective view derived from HRSC stereo images (orbit 2091).

to define areas of cold-climate environments where permafrost might be present although it is not a prerequisite (e.g., Washburn, 1979). Permafrost and ground ice produce a wide variety of landforms on Earth (e.g., Jahn, 1975; Embleton and King, 1975; Washburn, 1979; French, 1996; Yershov, 1998).

In the Viking era, permafrost and ground-ice-related landforms on Mars were reported more frequently than glacial ones. Early reports noted the morphological similarity of many Martian landforms to terrestrial analogs of the periglacial zone (Carr and Schaber, 1977; Rossbacher and Judson, 1981; Lucchitta, 1981). The advent of higher-resolution images and detailed global topography obtained through the MGS spacecraft (Albee *et al.*, 1998) spurred a large number of studies analyzing landforms on Mars reminiscent of terrestrial periglacial landforms.

10.11.3.5.2.(i) Lobate debris aprons, lineated valley fill, and concentric crater fill Debris flows that display similarities to terrestrial rock glaciers were first described in detail by Squyres (1978, 1979). He subdivided them on the basis of their morphology into lobate debris aprons (Figure 9), lineated valley fill, and concentric crater fill. Lobate debris aprons are characterized by a convex-upward topography and a steep and distinct flow front, strongly suggesting plastic deformation and flow. They can also show lineations on their surfaces, which can be both parallel and perpendicular to the flow direction. Where lobate debris aprons are confined by valley walls, their surfaces are deformed into



Figure 9 Lobate debris apron surrounding a mountain in the southern mid-latitudes ($\sim 41^\circ$ S, 103° E) on Mars. The morphology of the apron suggests viscous flow. Such deposits have been compared to terrestrial rock glaciers. Perspective false-color view of HRSC image 0451. The diameter of the debris apron is about 50 km (North is down, toward the viewer).

compressional lineations parallel to the confining walls. Material with this specific surface texture is called lineated valley fill (Figures 10(a) and 10(b)). Lobate debris aprons and lineated valley fill are best developed where mesas and valleys with steep walls are found, in particular at the steep scarps of the dichotomy boundary. Concentric crater fill is found in the interior of many craters, where ridges and troughs show a concentric pattern that can also show lobate flow fronts toward the crater center. All these types of landforms show evidence for plastic deformation and resemble terrestrial rock glaciers (Squyres, 1978, 1979; Lucchitta, 1984; Squyres *et al.*, 1989), which are mixtures of rock and ice (Wahrhaftig and Cox, 1959; Barsch, 1996). Whalley and Azizi (2003) give a comprehensive review on the terminology and problems related to the identification of rock glaciers on Mars.

Several hypotheses try to explain the origin of the ice in these debris flows. Water ice could have formed by direct condensation of ice from the atmosphere (Squyres, 1978) or by snow precipitation (Squyres, 1989). It could also accumulate by water vapor diffusion down into the regolith and subsequent condensation (Mellon and Jakosky, 1995). Finally, groundwater may seep into debris and create interstitial ice (Lucchitta, 1984; Squyres, 1989; Mangold and Allemand, 2001). The elastic particles in the lobate debris aprons might come from rock falls that accumulated at the base of scarps (Squyres, 1978; Colaprete and Jakosky, 1998) or, alternatively, from landslides (Lucchitta, 1984; Mangold and Allemand, 2001).

It has been demonstrated by Squyres (1978) on the basis of photoclinometry and by Mangold and Allemand (2001) and Li *et al.* (2005) on the basis of MOLA topographic profiles that the cross-sectional shape of lobate debris aprons can be approximated by the flow law of polycrystalline ice (Glen, 1955) and the flow relation of ice (Vialov, 1958; Paterson, 1984). Colaprete and Jakosky (1998) modeled flow of ice under Martian conditions and found that flow rates are very low. To create lobate debris aprons of the observed size, they show that temperatures 2040 K higher than present average mid-latitude temperatures (~ 210 K), ice contents exceeding 80%, and net accumulation rates of ≥ 1 cm per year are required.

Lobate debris aprons, lineated valley fill, and concentric crater fill are young landforms. Crater counts yielded low crater densities, and absolute ages of less than 100 Ma have been derived (e.g., Squyres, 1978; Mangold, 2003; Berman *et al.*, 2003; Head *et al.*, 2005; Li *et al.*, 2005). However, morphometric evidence for

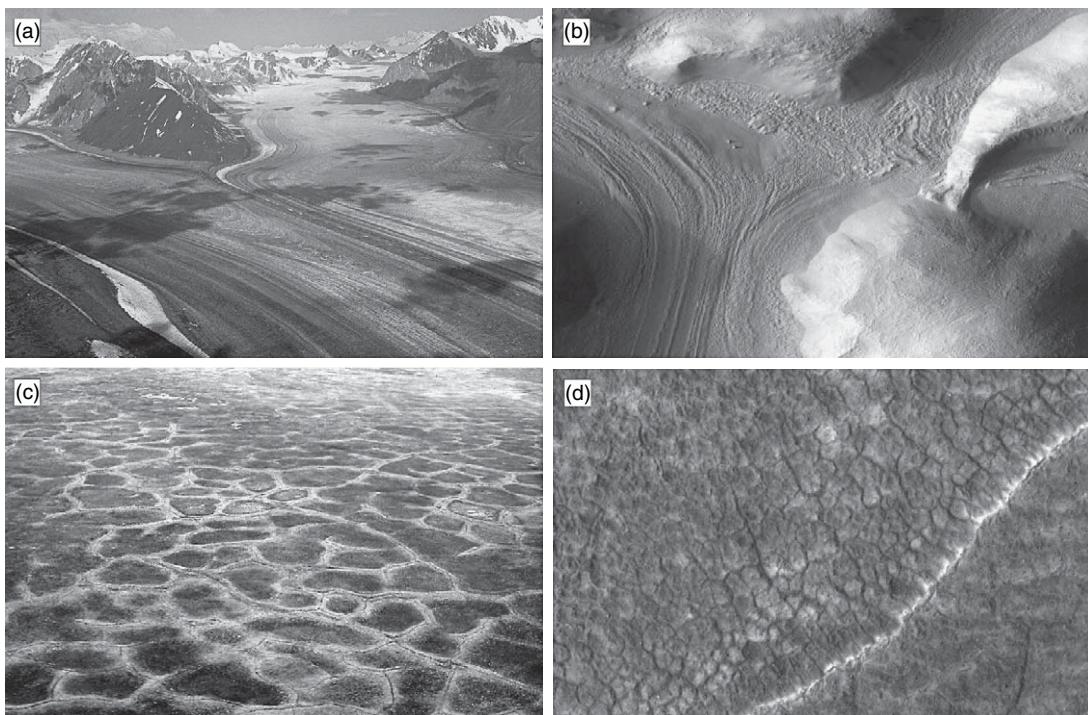


Figure 10 Comparisons between cold-climate landforms on Earth (left) and Mars (right). (a) Glaciers fusing together in the Yukon region (Canada; image courtesy Marli Miller, University of Oregon; image source: Earth Science World Image Bank). (b) Convergent flow features on Mars, probably the result of the flow of rock glacier-like material (MOC image SP245006). (c) Ice-wedge polygons in terrestrial permafrost terrain (Yukon, Canada; image courtesy Agriculture and Agri-Food Canada). (d) Patterned ground (polygons) on Mars. The diameter of the smaller polygons is less than 10 m and comparable to the size of ice-wedge polygons on Earth (HiRISE image TRA_000856_2265).

former lobate debris that were flooded by lava flows or debris flows (e.g., lahars) more than 1 Ga ago suggests that much older rock glaciers existed equatorwards of 30°, indicating a paleoclimate allowing ground ice to be stable at that latitudes (Hauber *et al.*, 2007).

10.11.3.5.2.(ii) Terrain softening At many locations, the Martian topography exhibits a rounded and subdued appearance when compared to undegraded terrain. The effect can be observed in particular at impact craters, which have rounded rims and convex-upward inner wall slopes. This phenomenon was termed ‘terrain softening’ by Squyres and Carr (1986) and mapped on a global scale. They found that this topographic softening affects terrain in the 30°–60° latitude belts. It affects all sorts of topographic features (Squyres *et al.*, 1989) and was attributed to the viscous relaxation of the topography due to the creep of a deformable and ice-rich regolith. While the mapping of Squyres and Carr (1986) was done qualitatively on the basis of Viking

Orbiter images, Jankowski and Squyres (1992) quantified the effect of terrain softening by examining the resulting topography by means of photoclinometry. Theoretical modeling suggests that the thickness of the deformable layer is limited to about 1 km (Jankowski and Squyres, 1992). The effect of viscous creep facilitated by ground ice was also suggested to explain the fact that rocky debris slopes are well below the angle of repose at higher latitudes (Perron *et al.*, 2003).

It has to be noted, however, that mantling deposits can also subdue the topographic signature of the surface. Soderblom *et al.* (1973) observed a mantling deposit with a similar extent than the terrain softening mapped by Squyres and Carr (1986). They attributed it to dust deposits from the polar regions. The effects of terrain softening and mantling can be discriminated because mantling processes tend to leave sharp topographic crests like crater rims unaffected, whereas terrain softening affects them strongly. In addition, mantling deposits remove

small topographic features, while relaxation processes leave them due to insufficient stresses (Squyres *et al.*, 1992).

MOLA data revealed that the global surface roughness displays a latitudinal trend (Kreslavsky *et al.*, 2000): at latitudes $>30\text{--}50^\circ$ in both hemispheres, Mars seems to be smoother at short baselengths than at lower latitudes. This observation is in agreement with the mapping of Squyres and Head (1986). Kreslavsky and Head (2000) suggest that this trend might be the result of a mantling deposit several meters high in high-latitude areas. Independently, it was found that images taken through the MOC show a very particular texture at certain locations. They reveal a dissected surficial layer or mantle, which suggests the degradation of an originally volatile-rich material (Mustard *et al.*, 2001; Malin and Edgett, 2001; Carr, 2001). Polewards of about 70° , the entire surface appears to be mantled, so that the surface looks very smooth and lacks sharp topographic gradients.

Mustard *et al.* (2001) suggested that the mantle was deposited as airfall dust, cemented by atmospherically derived ice. There are very few fresh impact craters superimposed on this surface, suggesting that it formed and degraded very recently. Global mapping showed that the mantle seems to be degraded in two latitude belts between 30° and 60° in both latitudes (Mustard *et al.*, 2001). These authors explicitly consider the idea of modern Martian ice ages operating at Milankovitch timescales, triggered by changes in insolation driven by variations in Mars' orbital parameters.

On a global scale, Milliken *et al.* (2003) mapped the occurrence of young gullies (Malin and Edgett, 2000c) (see Section 10.11.3.8), viscous flow features (e.g., Squyres and Carr, 1986), and the dissected terrain described by Mustard *et al.* (2001). Milliken *et al.* (2003) found a latitudinal concentration of these landforms at mid-latitudes with a spatial clustering at latitudes of $\sim 40^\circ$ in both hemispheres. They suggest that this correlation might be caused by the deposition, deformation, and removal of ice-rich materials at mid-latitudes in the Amazonian.

10.11.3.5.2.(iii) Thermokarst Thermokarst landforms develop as a result from the thawing of ice-rich permafrost or the melting of massive ice. Typical examples are thermokarst lakes, alases (large depression of the ground surface produced by thawing of a large area of very thick and exceedingly ice-rich permafrost), and thermokarst mounds (hummocks remaining after melting of the ice wedges surrounding

an ice-wedge polygon) (van Everdingen, 2005). On Mars, irregularly shaped depressions mainly in the outflow channels (Theilig and Greeley, 1979; Costard and Baker, 2001) and in the northern plains (Costard and Kargel, 1995; Bun Tseung and Soare, 2006; Morgenstern *et al.*, 2006) have been interpreted as thermokarst features.

A fluvial-induced mechanism of thermokarst on Mars has been proposed by Costard *et al.* (1999). It is based on analogy to huge terrestrial outflow channel-like streams in Siberia during snow-melt in spring, where the high water discharge transfers heat from the flow of water to the frozen ground and triggers ground thawing and thermal erosion. Costard *et al.* (1999) suggest that relatively warm floods on Mars might have enlarged outflow channels by a combination of thermal and mechanical erosion along frozen river banks.

10.11.3.5.2.(iv) Patterned ground A variety of landforms on Mars resemble terrestrial ice-wedging features. On Earth, this category of landforms includes polygons, circles, stripes, and nets (e.g., Washburn, 1979; French, 1996). Polygons, a common surface phenomenon in terrestrial permafrost areas (e.g., Lachenbruch, 1962), are the most frequently observed periglacial-like features on Mars. Viking Orbiter images showed several examples, however most of them had sizes that were an order of magnitude larger than their possible terrestrial analogs (Rossbacher and Judson, 1981). For example, giant polygons in Utopia Planitia with diameters of 20 km are probably too large to have originated by thermal cracking (Pechmann, 1980) and might instead be the product of fracturing of a sedimentary or volcanic layer due to cooling or dessication shrinkage plus differential compaction over a rough buried surface (McGill, 1986) or the result of tectonic uplift (Hiesinger and Head, 2000).

Possible periglacial landforms could be described in more detail on the basis of post-Viking data with a much higher spatial resolution. Small-scale polygons (**Figures 10(c) and 10(d)**) could now be analyzed morphometrically, and MOC images revealed that their size distribution is often much more like that of terrestrial ice-wedge polygons than it was thought before (Seibert and Kargel, 2001; Kuzmin and Zabalueva, 2003; Yoshikawa, 2003). Recent images with a resolution of less than 1 m per pixel, obtained by the HiRISE camera on board the MRO (McEwen *et al.*, in press), show that the diameters of polygons in Utopia Planitia are in the order of 10 m, which

corresponds exactly to the diameters of their terrestrial counterparts. Global-scale mapping using MOC images showed that the polygons are concentrated in two latitudinal belts poleward of 55° (Mangold, 2005) and correlate to the distribution of ground ice detected by the neutron spectrometer aboard Mars Odyssey (Mangold *et al.*, 2004). However, the absence of polygonal fracture patterns in the North Polar area and a high abundance of these landforms near the South Polar Cap suggests formation processes connected to seasonal deposition of CO₂ ice. Seasonal changes in polygon patterns were observed by van Gasselt *et al.* (2005) on the basis of multitemporal MOC images of a trough on the South Polar Cap. They ascribe these changes to contraction-crack processes in a thin layer of water ice beneath the seasonal CO₂ ice cover.

10.11.3.5.2.(v) Pingos and frost mounds Pingos on Earth are perennial frost mounds consisting of a core of massive ice, produced primarily by injection of water (Porsild, 1938; Mackay, 1973, 1979; Washburn, 1979). The identification of pingos in Viking Orbiter images has been ambiguous due to the relatively low image resolution (Rossbacher and Judson, 1981; Lucchitta, 1981, 1985). Higher-resolution MOC images allow a closer inspection of candidate topographic features, and several authors cite evidence for pingos at diverse locations like Gusev crater (Cabrol *et al.*, 2000), Athabasca Valles (Burr *et al.*, 2005), and the Cerberus Plains (Page and Murray, 2006), all at equatorial latitudes, and in the mid-latitude region of Utopia Planitia (Soare *et al.*, 2005).

10.11.3.6 Polar Caps

The inventory and the seasonal change of the water and carbon dioxide reservoir in the polar regions of Mars play an important role in understanding the evolution of the planet and its current state.

Viking data showed that the residual polar caps consist of H₂O ice. The North Polar Cap loses a seasonal coverage of CO₂ ice annually and the residual consist of H₂O ice (Kieffer *et al.*, 1976; Farmer *et al.*, 1976) The South Polar Cap retains in part their CO₂ coverage (Kieffer, 1979; Paige *et al.*, 1990). This survival of the CO₂ ice veneer on the South Polar was explained with a lower concentration of dust and therefore higher albedo (James *et al.*, 1992). However, it was unknown if the residual CO₂ ice is underlain by H₂O ice (Thomas *et al.*, 1992) The layered deposits

in both polar regions have been assumed to consist of H₂O ice (Cutts, 1973; Herkenhoff *et al.*, 1990).

Except for some spatially unresolved 10–50 μm IRIS spectra (Paige *et al.*, 1990) and TES spectra (Kieffer *et al.*, 2000), the spectral characteristics of the residual ice caps had not been published from any mission before Mars Express. The OMEGA (Bibring *et al.*, 2004a) and PFS (Hansen *et al.*, 2005) instruments on Mars Express independently detected and mapped spectral signatures of water ice on the Southern Polar cap of Mars. As expected OMEGA identified the 1.5 and 2.0 μm spectral signatures of water ice over extended parts of the polar cap. In the bright region the water ice is seen as a small admixture to CO₂ ice, whereas at the edge of the bright cap pure water ice with a dust admixture was seen. Surprisingly OMEGA detected also extended deposits of water ice in large areas tens of kilometres away from the polar cap. PFS – using a much higher spectral resolution but with a limited spatial coverage – observed a similar distribution of ice; however, there is a significant disagreement in the amount of water ice (Hansen *et al.*, 2005). While OMEGA measurements indicate a H₂O/CO₂ ratio of 15%, the PFS measurements indicate not more than several tens of parts per million by mass. The disagreement is most likely due to different modeling assumptions. While Bibring *et al.* (2004a) assume a molecular mixture of H₂O and CO₂ the analysis by Hansen *et al.* (2005) is based on a discrete condensation and therefore a heterogeneous mixture. Due to the spectral range used by the instruments they are inherently limited to probing the upper few millimeters of the surface and allow no conclusion about the bulk composition.

Studies of the evolution of the Northern Polar Cap indicate that observed brightness variations are due to changes in the grain size of the water ice (Langevin *et al.*, 2005b). In early summer seasonal water ice frost with grain sizes less than 100 μm sublimates and larger grains (~1 mm) of the permanent ice cap dominate the optical near-infrared spectral properties of the water ice.

Global neutron maps of Mars in various spectral ranges obtained by the HEND instrument on MGS allowed to constrain the content of water ice and adsorbed and bound water in the upper 1–2 m of the Martian surface. The data returned by this instrument showed huge regions of permafrost present both in the north of 55° N and south of 60° S. As discussed above the measurement of one energy range of neutron flux alone is not sufficient to derive

the water-ice content and distribution. However, combining data from neutrons with different energies and 2.2 MeV flux (Boynton *et al.*, 2002; Mitrofanov *et al.*, 2002; Feldman *et al.* 2002) indicates that a soil model consisting of a layer with water ice overlain by a layer of drier soil is in best agreement with the available data. The lower ice-rich layer contains between 50 and 75 wt.% water, with a higher abundance of water ice in shallow subsurface at northern polar latitudes than that in the south (Mitrofanov *et al.*, 2003). These ground-ice deposits extend over approximately 15% of the surface of Mars and are therefore a substantial water reservoir.

Previous compositional and stratigraphic interpretations of the polar cap deposits have been based on imaging, spectral, thermal, and topographic measurements. With the MARSIS radar on Mars Express it was possible to study the internal structure of the caps for the first time. In an early MARSIS orbit, 1855, the Northern Polar layered deposits were briefly observed in the longitude range from 10° to 40° E (Picardi *et al.*, 2005). The radargrams show the surface reflection splitting into a pair of strong reflectors as the ground track passes from the northern plains onto the layered deposits (Figure 11). An analysis of the signal taking into account the effect of surface topography as a source for clutter indicates that the second reflector is below the surface. The derived dielectric properties for the overlaying material indicate a fairly pure water ice with $\sim 2\%$ impurities and a bulk temperature of 240 K.

The delay time of the second reflection allows estimating the thickness of the overlaying material. The results indicate that basal reflector is approximately level with the surrounding plains. This excludes a significant downward deflection of the plains due to the load of the polar cap. A deflection of less than 500 m as indicated by the accuracy of the MARSIS measurements would put a lower limit of at least 150 km on the elastic thickness of the lithosphere (Johnson *et al.*, 2000).

10.11.3.7 Large Debris Flows

A large variety of gravity-driven flow or mass-movement processes that involve water and sediments have been termed debris flows (see reviews by Coussot and Meunier (1996) and Iverson (1997)). Several landforms on Mars have been interpreted as the products of erosion and sedimentation by debris flows. While the small-scale gullies and the rock glacier-like features are discussed elsewhere in this

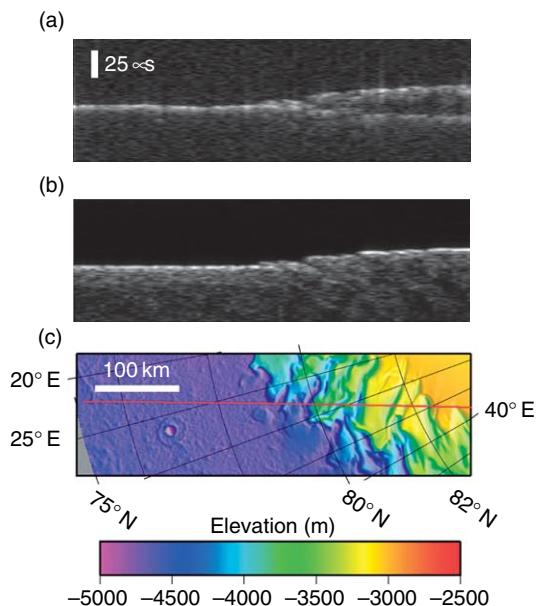


Figure 11 (a) MARSIS data in radargram format for orbit 1855 as it crossed the margin of the NPLD. (b) Simulated MARSIS data if echoes are only from the surface (nadir and off-nadir clutter). (c) MOLA topography along the ground track (red line); elevation is relative to mean planetary radius. MARSIS data at 5 MHz show a split of the strong return into two as the ground track reaches the NPLD (higher terrain to the right). Maximum time delay to the second reflector is 21 μ s, equivalent to 1.8-km depth in water ice. From Picardi G, Plaut JJ, Biccarri D, *et al.* (2005) Radar soundings of the subsurface of Mars. *Science* 310: 1925–1928.

paper (see Sections 10.11.3.8 and 10.11.3.3), here we give a short overview of possible large-scale debris flow processes on Mars. Debris flows were invoked several times to explain the complex assemblages of collapsed surface regions and the huge outflow channels. For example, a debris flow mechanism was proposed to account for the formation of the chaotic terrains and the outflow channels debouching into Chryse Planitia (Nummedal and Prior, 1981). Huge smooth deposits on the floor of some outflow channels and extending into Chryse Planitia have been interpreted to be debris flows (Tanaka, 1997, 1999).

Large flow-like deposits north-west of the Elysium volcanic rise might have been emplaced as lahars (Christiansen, 1989; Russell and Head, 2003). Lahars are debris flows originating at volcanoes, often triggered by the melting of ice (glaciers, ground ice) by volcanic heat (e.g., Vallance, 2000). Water was also proposed to be a constituent in Martian landslides, mainly in Valles Marineris (Lucchitta, 1979, 1987). It has to be noted, however, that the presence

of water in Martian landslides is still under debate, and a dry origin for the landslides in the Valles Marineris also seems possible (McEwen *et al.*, 1989), a view that was recently confirmed by lab-scale dry granular flow experiments (Lajeunesse *et al.*, 2006) and comparative studies with landslides on Earth (Soukhovitskaya and Manga, 2006).

It should be mentioned that interpreting debris flow features and deriving implications on their water content and emplacement mechanisms is not without problems on Mars. The strong and possible long-term modification of landforms on Mars by cold-climate processes often hinders the identification of their original nature (Van Gasselt *et al.*, in press).

10.11.3.8 Gullies

The discovery of gullies on Mars was one of the most exciting findings of the MGS mission because most gullies are pristine and not superposed by impact craters, which is a sign of their formation in recent geological timescales. Because of their small scale they were first detected in high-resolution MOC images (Malin and Edgett, 2000c). The morphology of these features indicates the involvement of a liquid, most probably water, in their formation. As a consequence the gullies are important for the Martian history because they may attest to very recent hydrological activity on Mars.

10.11.3.8.1 Morphology

Gullies on Mars resemble terrestrial features formed by mass-wasting processes of a flowing mixture of debris and water (debris flows) (Figure 12). On Earth, debris flows commonly occur in semiarid, alpine and arctic regions and are frequently triggered by intense rainfall or sudden thawing of frozen ground (Costa, 1984). In their majority they are caused by liquefaction of water-saturated debris on talus slopes. Like on Earth, the Martian features show three typical morphological attributes of the result of a debris flow which can be divided into (1) head alcove (source region), (2) channel (transport zone), and (3) debris apron (deposition zone) (van Steijn, 1989; Malin and Edgett, 2000c). Gully sizes range from lengths of a few hundred meters to some kilometers and widths of some meters to several tens of meters identical to the terrestrial features. Morphological details of the gullies suggest that a fluid medium is responsible for their formation (Malin and Edgett, 2000c). The source regions can show contributory anastomosing patterns which merge into a main channel. The main channels are often accompanied by secondary channels which both frequently show lateral ridges (levees), sinuous paths, branched reaches, deep incision, and are streamlining around obstacles. The depositional aprons often display distinct lobate swells or digitate flows. In general, the gullies show a

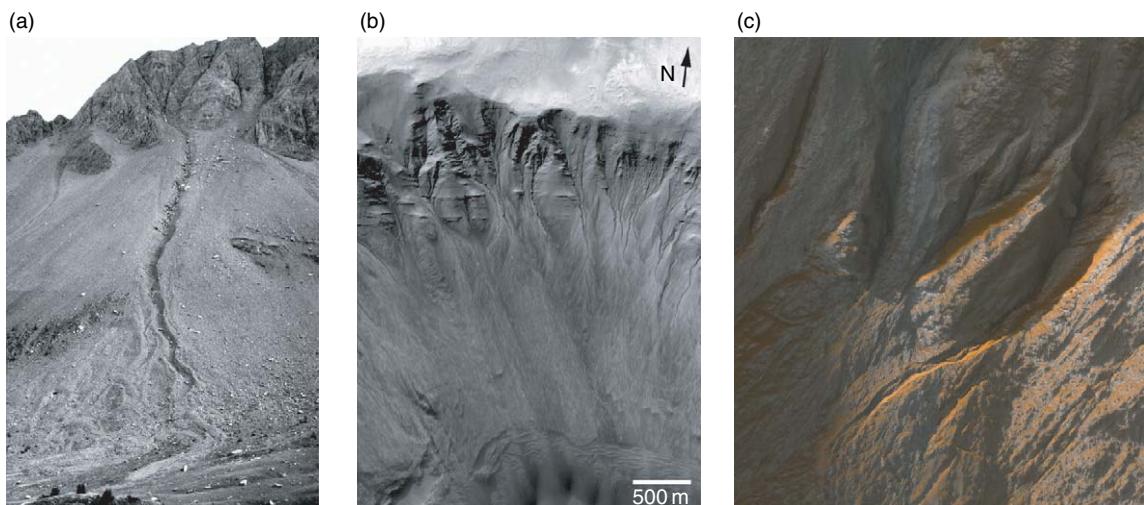


Figure 12 (a) Terrestrial debris flow in the French Alps, south of Barcelonnette (D. Reiss, Aug 2003). (b) Gullies on a slope of a crater located in the Newton basin (MOC-image E1600043 at 201.8° E and 41.8° S, 3 m per pixel). (c) Detail color view of gullies on the slope of an unnamed crater in Terra Sirenum. Bluish areas are most likely composed of water frost (HiRISE press release image TRA_000878_1410_RGB at 223.7° E and 38.9° S, 50 cm/pixel, image width is 250 m).

broad variety of morphologies and little research has been done so far to distinguish them among each other.

10.11.3.8.2 Distribution

Gullies on Mars are tied to steep terrains. Most of them occur on the walls of craters, less frequently of troughs, valleys, and pits. Other locations include slopes of crater central peaks, isolated knobs, and dunes. They occur regionally concentrated in specific regions like single craters and polar pits (Malin and Edgett, 2000c). The global distribution is limited to mid- and high-latitudes poleward of 30° in both hemispheres, but most frequently in the 30° – 50° latitude bands. The only relevant concentration of gullies equatorward of 30° latitude occurs on the north wall of Nirgal Vallis at 27° latitude. Gullies are much more common in the Southern Hemisphere. This asymmetry is possibly due to the less cratered and smooth terrain in the Northern Hemisphere in contrast to the heavily cratered southern highlands with a high amount of steep slopes (Malin and Edgett, 2001).

First global studies showed a preferred poleward-facing slope direction of gullies (Malin and Edgett, 2000c). Further studies with more data show a more complex view: while Edgett *et al.* (2003) found no preferred orientation of gullies in both hemispheres, others found that in the Northern Hemisphere they occur preferentially on equator-facing slopes (Bridges and Lackner, 2005; Heldmann *et al.*, 2005) and that in the Southern Hemisphere gullies predominate in the mid-latitudes on pole-facing slopes (Heldmann and Mellon, 2004; Balme *et al.*, 2006; Dickson *et al.*, in press) with a change to a general occurrence on all slope directions towards higher latitudes (Heldmann and Mellon, 2004; Balme *et al.*, 2006). The most recent analyses seem to confirm a latitude-dependent change in the orientation of gullies in both hemispheres.

10.11.3.8.3 Formation

The formation mechanism for gullies is controversially debated since their discovery. The morphology of the gullies suggests that they are formed by a mass-wasting process behaving like a fluid. The fluid medium as well as the source of it is still discussed. Hoffman (2000b, 2002) and Musselwhite *et al.* (2001) suggest that carbon dioxide from surface snow or the subsurface can carve the channels by gas-lubricated flows with entrainment of clastic debris. Additionally, Jöns (2002) suggests that the decomposition of CO₂-

hydrates can deliver sufficient amounts of liquid water for the erosional processes. These theories are unlikely because neither CO₂ nor CO₂-hydrates can accumulate in significant quantities in the subsurface and the morphology of possible gas-supported flows is not consistent with the observed gully morphology (Stewart and Nimmo, 2002). Granular flows are suggested by Treiman (2003) and Shinbrot *et al.* (2004), but dry mass movements cannot form the observed the characteristic morphologies. However, there is a broad variety of different morphologies and some Martian features may resemble dry mass movement features on the Moon (Bart, 2007). The most likely medium to explain the gully morphology is liquid water. Two main theories depending on the source of water exist. The water is suggested to be released from the subsurface or deposited from the atmosphere and melted by insolation. Malin and Edgett (2000c) propose a model in which a shallow subsurface reservoir of liquid water is protected against evaporation by an ice barrier. Sudden release of liquid water and subsequent debris flows occur when the icy plug is ruptured (Malin and Edgett, 2000c). Gilmore and Phillips (2002) favor melting of near-surface ice which percolates through a permeable surface layer and subsequent seepage at the base of an impermeable layer in the subsurface. Others suggest that geothermal heating melts subsurface ice (Gaidos, 2001; Hartmann, 2001; Mellon and Phillips, 2001). Atmospheric-dependent models suggest that the water ice is deposited out of the atmosphere and melted by insolation during periods of high-obliquity phases of Mars. Costard *et al.* (2002) suggest accumulation of ice within the regolith and melting of this near-surface ice by insolation during periods of high obliquity. In this model the near-surface ice is protected against evaporation by seasonally deposited CO₂-frost. Lee *et al.* (2001) and Christensen (2003) propose a model in which water ice is deposited as snow during high-obliquity periods and melted at low obliquity. Here, melting occurs beneath the snow cover sheltering the liquid water against evaporation (Christensen, 2003).

The limited global distribution of gullies in mid- and high-latitudes and the preferred orientation on poleward-facing slopes coincides with both water-based main theories of gully formation. Water might have been stable in these latitude regions as well as most likely in sheltered regions on poleward-facing slopes. Higher temperatures near the equator and on equator-facing slopes would quickly evaporate emergent surface water or groundwater (Costard

et al., 2002; Malin and Edgett, 2000c). Geothermal heating is unlikely because there is no correlation to volcanic regions on Mars (Malin and Edgett, 2000c). The observation of gullies on dunes (Costard *et al.*, 2002; Reiss and Jaumann, 2003; Mangold *et al.*, 2003) and impact crater central peaks (Baker, 2001) question the theory of subsurface seepage to form the gullies, because the involvement of an aquifer is unlikely. The atmospheric model is consistent with other recent morphological observations of a deposited water-ice mantling and viscous flow features in the mid-latitudes (Mustard *et al.*, 2001; Milliken *et al.*, 2003). The formation of an ice-rich mantle, viscous flow features, and gullies are possibly all related to recent ice ages within the last few million years on Mars. According to this theory water ice was removed from the poles and transported and deposited as a dust/ice mantle to mid-latitudes during higher obliquities, which desiccated at lower obliquities (Head *et al.*, 2003a). There are some inconsistencies in the effort of different analyses to correlate the gully formation to times of higher obliquities. Head and Marchant (2003a) predict gully formation at obliquities higher than $>30^\circ$ and Costard *et al.* (2002) at obliquities higher than $>35^\circ$, whereas Mellon and Phillips (2001) do not predict any melting of near surface ice at higher obliquities. Additionally, all the morphologic and obliquity model predictions do not coincide. The model of Christensen (2003) depends on a slow gully formation (thousands of years), which is in contrast to the rapid and episodic formation, which was suggested on the basis of morphology (Malin and Edgett, 2000c; Heldmann *et al.*, 2005). Further work is needed to constrain the formation of Martian gullies and it is crucial to note that not all gullies must have been formed exclusively from the same process. As on Earth, different processes are able to form the same morphologic features, although water remains the most likely medium involved in the gully-forming process.

10.11.3.8.4 Age

The gullies are suggested to be formed geologically recently. Estimates that the gullies could be less than 1 Ma years old are based on their pristine appearance, their stratigraphic relationships to young surface features, their lack of superimposed impact craters, and their distinct albedo to the surroundings, indicating limited dust cover (Malin and Edgett, 2000c). Other age estimations assume that the gullies were formed under different

climatic conditions than today and correlate their activity to obliquity variations in the last few million years (Lee *et al.*, 2001; Costard *et al.* 2002, Christensen, 2003; Head *et al.*, 2003a). Crater counts on dunes superposed by gully deposits in Nirgal Vallis indicate that the formation processes were active within the last 3 Ma, possibly within the last 300 000 years (Reiss *et al.*, 2004). All these ages derived from morphologic observations, modeling, and crater counts are relatively consistent and imply that the gullies might be younger than a few million years. However, they are based on rough estimates or are regionally limited. Recent observations show newly formed gully deposits (Figure 13) and their morphologic characteristics suggest that liquid water flowed on the Martian surface within the past 2 Martian years (Malin *et al.*, 2006). Although most of the gullies are morphologically pristine, there is observational evidence of older gullies which are degraded and cratered. Degradation of gullies can be seen as filled channels, eroded aprons, and gullies cut by fissures. Others are covered by small impact craters, which show additionally degraded morphologies. These older examples indicate that not all gullies are young and the gully-forming processes occurred at least episodically over a longer timescale.

10.11.4 Water in Mantle and Crust

Looking at the terrestrial planets today, Mars is the only planet on which surface water has been detected directly in the form of ice at the polar caps (Bibring *et al.*, 2004a; Hansen *et al.*, 2005). There is also evidence for large amounts of ground ice in the circumpolar areas (Feldman *et al.*, 2004). Recent radar observations by the MARSIS instrument on Mars Express (Picardi *et al.*, 2005) and theoretical modeling indicate the existence of ground-ice deposits extending to mid- and low-latitudes (Helbert *et al.*, 2005). For Mercury (Harmon and Slade, 1992; Butler *et al.*, 1993) (as well as for the Moon (Feldman *et al.*, 1998; Nozette *et al.*, 1996)) there are indications for the existence of ice deposits in the regolith in permanently shadowed areas in the polar regions. These deposits are still subject to debate and await confirmation by future missions. The present-day conditions clearly exclude the existence of liquid water or ice on the surface of Venus.

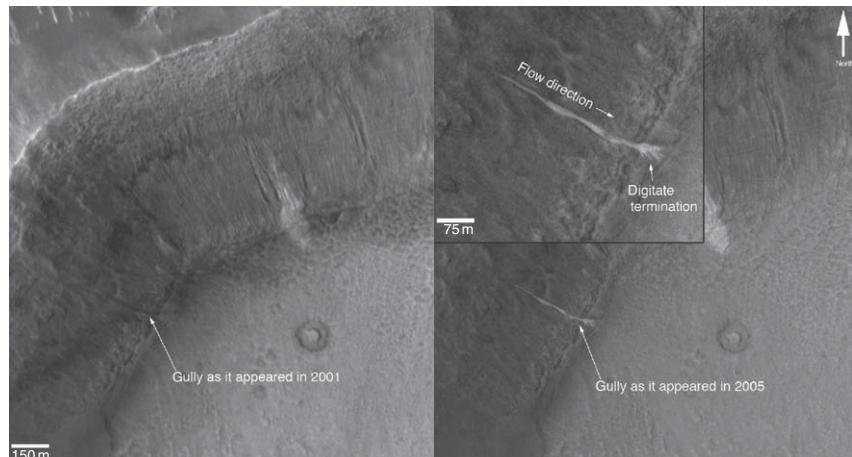


Figure 13 Newly formed gully deposits between Dec 2001 (left MOC image, E1103412) and Apr 2005 (right MOC image, S0902603 and S1001184). The digitate distal and marginal branches of the light-toned deposits (inset of right image), which divert around obstacles suggest a very fluid material (Malin et al., 2006) (NASA/JPL/Malin Space Science Systems).

10.11.4.1 Regolith

10.11.4.1.1 Moon and Mercury

The rotation axis of the Moon has an inclination of only 1.6° resulting in areas near the poles, primarily the floors and lower interior walls of impact craters, that are permanently shadowed from the Sun. It was suggested already very early (Watson et al., 1961; Arnold, 1979) that temperatures within these permanently shadowed regions near the lunar poles were sufficiently low that water would be stable to sublimation for the lifetime of the Moon. Meteoroids and comets were proposed as source for the water. A number of follow-up studies refined this idea (Butler et al., 1993; Butler, 1997; Vasavada et al., 1999). The first tentative observation indicating possible ice at the lunar South Pole was reported by Clementine from bistatic radar observations of very low spatial resolution (Nozette et al., 1996). The 13-cm-wavelength telemetry system of the Clementine orbiter was used as a transmitter and one of the 70-m NASA/Deep Space Network antennas to receive the echo. The signal received from the pass over the South Pole showed a lower depolarization as the other passes indicate the presence of a low-loss volume scatterer, possibly water ice. Following analysis by Nozette et al. (2001) suggested a body of $\sim 10 \text{ km}^2$ of ice in the Earth-facing lower part of Shackleton crater. However, analysis of Arecibo radar data (Stacey et al., 1997) and a re-analysis of the Clementine data (Simpson and Tyler, 1999) have raised questions regarding this conclusion. An independent indication for possible

ice deposits at the poles of the Moon was provided by an analysis of Lunar Prospector (LP) neutron data. This data showed enhanced abundances of hydrogen, existing at both poles of the Moon (Feldman et al., 1998, 2000). The spatial resolution of the LP measurements was not sufficiently good to prove a one-to-one correspondence of the signal to the permanently shaded craters near both poles (Margot et al., 1999; Bussey et al., 1999). However, it was possible to show that an association was plausible. Using models of neutron emission from assumed water-ice deposits filling the largest three craters in the neighborhood of the South Pole a water-ice abundance of $1.5 \pm 0.8\%$ by mass was derived. The result is however not a proof for the existence of water ice, because the measurement of hydrogen is independent of its chemical association. Recent high spatial resolution Earth-based radar observations using the 13 cm radar-imaging with the Arecibo radar system showed no evidence for characteristic changes in the depolarization in Shackleton crater or anywhere else near the South Pole of the Moon (Campbell et al. 2006). These results in combination with the LP results imply a low likelihood for a thick ice deposit within the crater but rather favor ice being dispersed within the regolith.

The same arguments made for the Moon are valid for Mercury as well. Thermal model (Ingersoll et al., 1992; Paige et al., 1992) showed that ice can remain stable over geological timescales in permanently shadowed areas near the poles of Mercury. This was confirmed by Earth-based radar observation discovering probable ice deposits at the poles of Mercury

(Slade *et al.*, 1992; Harmon and Slade, 1992). The radar echoes from Mercury's poles were interpreted as reflections from water ice on the basis of their similarity to the unusual properties of radar echoes from the icy Galilean satellites of Jupiter (Campbell *et al.*, 1978). Higher spatial resolution radar images (see Figure 14), obtained by Harmon *et al.* (2001), confirmed all the original polar features and revealed many additional features, including several at

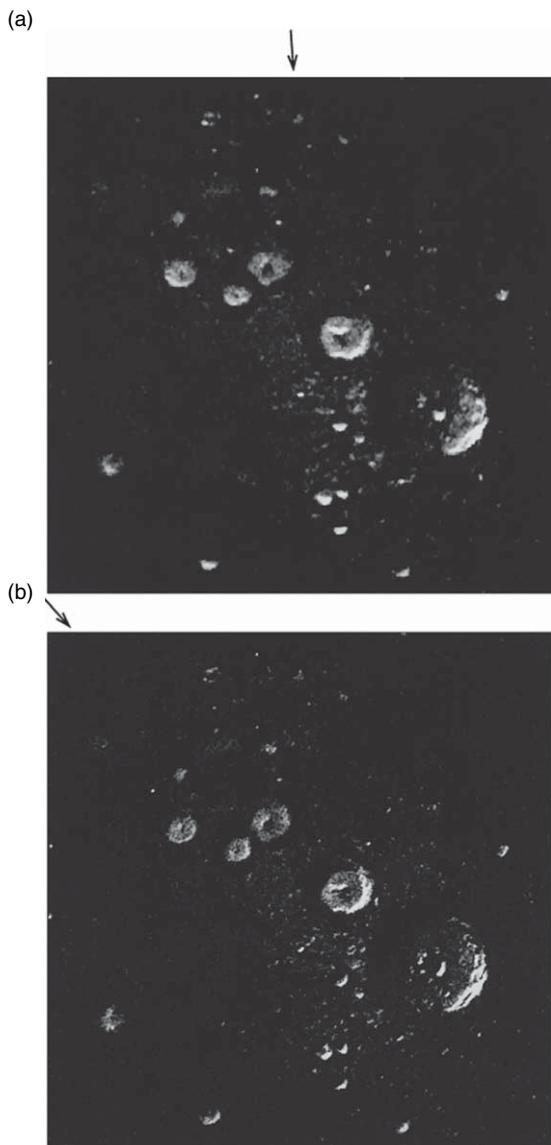


Figure 14 Radar images of the deposit close to the pole of Mercury. Lighter shades correspond to higher echo strength. The radar illumination directions are indicated (arrows). Surface resolution is 3 km in (a) and 1.5 km in (b). From Harmon JK, Perillat PJ, and Slade MA (2001) High-resolution radar imaging of Mercury's north pole. *Icarus* 149: 1–15.

latitudes as low as 72–75° N and several from craters less than 10 km in diameter. All of the new features located on the Mariner-imaged side of the planet could be matched with known craters or other shaded areas. Moses *et al.* (1999) showed by extrapolating the current terrestrial influx of interplanetary dust particles to that at Mercury, that continual micrometeoritic bombardment of Mercury over the last 3.5 billion years could have resulted in the delivery of $(3\text{--}60) \times 10^{13}$ kg of water ice to the permanently shaded regions at Mercury's poles. This would result in an average ice thickness of 0.8–20 m. Erosional processes like, for example, solar wind sputtering or micrometeorite bombardment would remove about half of this ice in the same time span. The resulting amount of ice is in general agreement with the proposed amount of ~2–20 m based on radar imaging if the additional influx by comets and asteroids is taken into account. Butler *et al.* (1993) discussed the notion that water originating away from the poles would behave as individual, randomly moving molecules, some of which could migrate to the poles and become cold-trapped there. An independent confirmation of the proposed ice deposits has to wait until the arrival of the NASA mission MESSENGER and the ESA mission BepiColombo. (McNutt *et al.*, 2004)

10.11.4.1.2 Mars

The stability of water-ice on the surface of Mars is governed by the balance between sublimation and deposition and mainly a function of the atmospheric temperature and water vapor abundance. For subsurface water-ice deposits in the regolith the situation is more complex, because the exchange rate with the atmosphere is controlled by the physical properties of the regolith. In principle two end-member situations can be identified. One extreme scenario is that the subsurface ice is disconnected from the atmosphere by an impermeable layer of material. The other extreme would be a free exchange of water vapor between the subsurface ice deposit and the atmosphere. While the regolith would protect the subsurface ice deposit from temperature variation, ground ice and atmospheric vapor would eventually reach equilibrium (Flasar and Goody, 1976; Mellon and Jakosky, 1993; Schorghofer and Aharonson, 2005). Between these two extremes there are a variety of other possibilities. Helbert *et al.* (2005) proposed for example that the retreating ground ice forms a sublimation till which limits the vapor transport to the atmosphere thereby stabilizing the ice.

The wide range of assumptions on the interaction between subsurface ice and the atmosphere lead to a disagreement about the stability of proposed ground-ice deposits especially at mid- and low-latitudes. The calculations by Farmer and Doms (1979) and Mellon and Jakosky (1993, 1995) show that water ice must be unstable at shallow depths for latitudes less than 65°. Schorghofer and Aharonson (2005) took into account seasonal variations of air moisture in their model and found that the ice can be stable under some conditions, in particular, at low thermal inertia. In the model by Helbert *et al.* (2005), subsurface ice is metastable at these depths: it disappears, but slowly in geologic scale. On Mars the variation of climate over the history of the planet plays an important role in the distribution and amount of ice in the regolith.

Impact cratering into ice-rich (and possibly fluid) target material is suggested to form a specific layered ejecta morphology on Mars termed Rampart crater, which is not observed on other terrestrial planets. Models suggest that the target material must contain at least 15–20% ice for the formation of a layered ejecta blanket (Woronow, 1981; Stewart *et al.*, 2001). The ejecta are characterized by ‘lobate’ or ‘fluidized’ like structures, which terminate in distal ramparts (**Figure 15**). A summary of different layered ejecta morphologies is given in Barlow *et al.* (2000). Although most researchers suggest impacts into volatile-rich target material for their formation, (e.g., Head and Roth, 1976; Carr *et al.*, 1977; Mouginis-Mark, 1979) atmospheric processes might also cause or contribute to it (e.g., Schultz and Gault, 1979; Schultz, 1992; Barnouin-Jha and Schultz, 1998). For a summary of models, morphologic constraints,

laboratory experiments, and discussion about the two formation theories see the reviews by Barlow (2005) and Kuzmin (2005).

Rampart crater are found in all geological units, and at all altitudes and latitudes (Allen, 1979; Mouginis-Mark, 1979; Kuzmin *et al.*, 1988; Costard, 1989; Barlow and Bradley, 1990). In a given area a certain minimum diameter exists for Rampart craters, called the onset diameter (Boyce and Witbeck, 1980; Kuzmin, 1980). Geographic mapping revealed that the onset diameter decreases towards higher latitudes (Mouginis-Mark, 1979; Kuzmin *et al.*, 1988; Costard, 1989). Rampart crater diameters in equatorial regions are typically 4–7 km versus 1–2 km at high latitudes (50° latitude) (Squyres *et al.*, 1992). Based on the physics of the impact process and using a relationship between the onset diameter and the crater’s excavation depth, Kuzmin *et al.* (1989) derived global depths of the boundary between a desiccated and ice-bearing regolith layer. The results indicate a possible boundary in ~300 to ~400 m near the equator and ~100 m at 50° latitudes (Kuzmin, 2005). Barlow *et al.* (2001) showed that the onset diameter varies regionally near the equator, indicating shallower ground-ice depths in these regions. Recently, Reiss *et al.* (2005) found regionally equatorial onset diameter of Rampart craters as small as 1 km, equal to those at high latitudes. However, the layered ejecta blankets are highly degraded, indicating a formation in the early history of Mars and possible shallow ground-ice depths at these times (Reiss *et al.*, 2005). Rampart craters may have formed over a significant time interval and the onset diameter records the ground-ice depth at the time of the impact (Squyres *et al.*, 1992). Possible changes with time might be detectable. Barlow (2004) analyzed characteristics of layered ejecta in equatorial regions and found no major variations in volatile concentrations at depths >1 km. Based on absolute age determinations by crater counts on the layered ejecta of Rampart craters in equatorial regions and in comparison to the onset diameter, Reiss *et al.* (2006) suggest a decrease in the ground-ice table at depths <1 km in these regions.



Figure 15 Perspective color view of a Rampart crater (diameter = 9.5 km) in southern Arabia Terra, Mars (HRSC orbit 2956_0000 at 16.5° E and 15.2° N).

10.11.4.2 Mineralogy

Minerals are the basic rock-forming constituents of planetary surfaces. As such, their investigation can provide valuable information about the aqueous history of a planet. The most interesting minerals in this respect are hydrated minerals, that is, minerals which include water (H_2O) or hydroxyl (OH) in their

mineral structures. Examples of hydrated minerals are iron hydroxides, carbonates, sulfates, phyllosilicates, and zeolites. The only terrestrial planets which have been found to have hydrated minerals at their surfaces are the Earth and Mars. Some asteroids might also possess OH and/or H₂O-bearing minerals (Rivkin *et al.*, 2002), for example, Vesta (Hasegawa *et al.*, 2003), but these are not further discussed here. On the other hand, there is no evidence for water-containing minerals on the Moon (Lucey *et al.*, 2006), and so far on Venus although here too little of the mineralogy is known to make a firm statement. It is not the scope of this paper to discuss hydrated minerals on Earth. Summaries of Martian surface mineralogy, including potential hydrated minerals, are given by Soderblom (1992), Roush *et al.* (1993), Bell (1996), and Bibring and Erard (2001). Bish *et al.* (2003) discuss the stability of hydrous minerals on Mars, Bishop (2005) gives a recent, but still pre-MER and pre-MEX overview of the properties of hydrated minerals on Mars, and a review of the several forms of hydration that can exist on the Martian surface is presented by Jouget *et al.* (2006).

10.11.4.2.1 Methods

Typically, the composition of Martian surface materials has been analyzed on the basis of Earth-based spectroscopic (e.g., McCord and Adams, 1969), *in situ* chemical (e.g., Clark *et al.*, 1977; Rieder *et al.*, 1997; Wänke *et al.*, 2001) and *in situ* magnetic (e.g., Hargraves *et al.*, 1977) data. Reflectance spectroscopy is a particularly powerful tool to infer the mineralogy of planetary surfaces (e.g., Burns, 1970), but only few instruments were operated in orbit around Mars. The near-infrared spectrometers IRS aboard Mariner 6 and 7 (Calvin, 1997) and imaging spectrometer for Mars (ISM) aboard Phobos-2 (Bibring *et al.*, 1989; Murchie *et al.*, 1993) detected a broad and deep 3 μm absorption band in each spectrum. The strength and shape of the 3 μm band is influenced by all forms of hydration. Unfortunately, the ISM acquired only a very limited number of observations before the Phobos-2 mission failed at an early stage. The most recent near-infrared imaging spectrometers in orbit around Mars are the OMEGA on board Mars Express (Bibring *et al.*, 2004b) and the compact reconnaissance imaging spectrometer for Mars (CRISM) on board the MRO (Murchie *et al.*, 2004).

TES on board the MGS Mission (Christensen *et al.*, 1992) and the THEMIS on board the Mars Odyssey Mission (Christensen *et al.*, 2004a) both

acquired (and acquire) spectra at longer wavelengths in the thermal emission range.

The MERs, Spirit, and Opportunity, have several instruments which can provide mineralogical information (Squyres *et al.*, 2003). The Panoramic Camera (Bell *et al.*, 2003) and the miniature thermal emission spectrometer (Christensen *et al.*, 2003) provide multi-spectral data and can still be considered to be ‘remote sensing’ experiments, despite their very close range to the observed targets. The Mössbauer spectrometer (Klingelhöfer *et al.*, 2003) and the alpha particle X-ray spectrometer (APXS) (Rieder *et al.*, 2003), which provides geochemical rather than mineralogical information, take their measurements in direct contact with the target materials.

10.11.4.2.2 Basalt and dust

Overall, the surface of Mars consists of a mixture of relatively pristine mafic materials (i.e., basalt consisting of pyroxenes and olivine) and highly oxidized secondary materials (Bell, 1996). Mössbauer and APXS data collected by the MERs show that the ubiquitous, oxidized bright red dust, which is responsible for the red color of Mars, consists of ferric oxides and includes nanocrystalline phases (e.g., Goetz *et al.*, 2005). It is an alteration product, probably of a primary material with basaltic chemistry, but the presence of olivine in the dust, which is quickly weathered in the presence of water, suggests that water did not play a significant role in its production (Goetz *et al.*, 2005). This is also confirmed by orbital observations with the OMEGA instrument. It found evidence for widely distributed anhydrous ferric oxides in high-albedo regions (Bibring *et al.*, 2006), which might consist of nanocrystalline red hematite (α -FeO) and maghemite (γ -FeO). Therefore, both rover and orbiter data point to the formation of the Martian dust that occurred under a dry climate.

While basaltic rocks and anhydrous ferric oxides (dust) dominate the Martian surface mineralogy, several types of water-related minerals have been identified so far in the uppermost surface materials on Mars, although at very isolated locations.

10.11.4.2.3 Hematite

Spectra taken with TES show evidence for gray, coarse-grained (>10 μm) crystalline hematite in three limited regions, Aram Chaos, Meridiani Planum, and, to a lesser degree, Valles Marineris (Figure 16) (Christensen *et al.*, 2000, 2001b).



Figure 16 Hematite in Meridiani Planum as mapped with the TES (e.g., Christensen *et al.*, 2000). The hematite-rich areas (marked in red) have probably experienced long-lasting aqueous processes and were selected as landing site for one of the MERs. Image: NASA Planetary Photojournal, PIA02056.

Hematite is not a hydrated mineral, but it is thought to have formed via aqueous processes. Although several mechanisms exist to form this variety of hematite, Christensen *et al.* (2001b) favor chemical precipitation from Fe-rich aqueous fluids under either ambient or hydrothermal conditions (see also Catling and Moore, 2003). Such a process would have required substantial amounts of liquid water. From a morphological perspective this appears plausible, since all three hematite deposits detected with TES are located in sedimentary settings. Actually, the hematite deposit in Meridiani Planum was considered to be such strong evidence for past surface or near-surface water that one of the MERs, Opportunity, was sent there to explore its details (Golombek *et al.*, 2003).

Indeed, the findings of the rover confirmed the presence of hematite in Meridiani Planum. Over the entire region, small millimeter-sized spherules (called ‘blueberries’ by the rover team) are embedded in the sandstones which form the layered outcrops in Meridiani Planum. These spherules were found to consist of hematite, based on the observations of the Mössbauer and Mini-TES instruments (e.g., Klingelhöfer *et al.*, 2004; Christensen *et al.*, 2004c). Their accumulation in topographic depressions as a result of weathering is responsible for the detection of hematite from orbit.

10.11.4.2.4 Phyllosilicates

Phyllosilicates form a large group of minerals that includes among others micas, smectites, and kaolinite. Their basic characteristic feature is the

layered structure, and their formation can take place at the surface (e.g., in soils or sediment) or subsurface (diagenesis, hydrothermal alteration) (e.g., Meunier, 2005). Some of them, for example, smectites, form as alteration products of volcanic material. Therefore, their presence on Mars has been considered possible and even likely on theoretical grounds (e.g., Bishop, 2005). Several studies suggested the existence of phyllosilicates on Mars based on *in situ* Viking Lander analyses (Toulmin *et al.*, 1977), the identification of smectites in laboratory analyses of shergottite, nakhlite, and chassigny (SNC) meteorites (e.g., Bridges *et al.*, 2001), and spectroscopic measurements (e.g., Clark *et al.*, 1990; Bishop *et al.*, 1995; Wyatt and McSween, 2002). However, the general opinion, confirmed by the detection of sulfates in orbiter and rover measurements, was that the Martian mineralogy was dominated by acid-sulphate processes, which prohibit the formation of clay minerals (see the comment by Newsom (2005)). Nevertheless, phyllosilicates were identified on Mars in near-infrared spectra taken with the OMEGA spectrometer (Bibring *et al.*, 2005). Clay minerals such as smectites were mapped in two major outcrops in the area of Mawrth Vallis and Nili Fossae, and in several smaller-sized areas (Poulet *et al.*, 2005). The phyllosilicate-rich materials were located in ancient (Noachian) regions, but their surfaces do not show large abundances of impact craters and, therefore, must have been exposed relatively recently.

Phyllosilicates have also been found in Gusev crater at the base of the Columbia Hills. Measurements of the Mössbauer and APXS spectrometers indicate the presence of members of the kaolinite, serpentine, chlorite, and sepichlorite groups (Wang *et al.*, 2006). The detection of Al-rich phyllosilicates suggests a mildly acidic past environment and an open hydrologic system with good drainage conditions (Wang *et al.*, 2006).

10.11.4.2.5 Sulfates

Sulfates are salts of sulfuric acid and often contain water molecules. They are particularly useful to constrain the climatic evolution, since some sulfates can be transformed to other sulfates that have a different water content. While this reaction is sometimes reversible, it is irreversible for others. As an example, kieserite ($\text{MgSO}_4 \cdot \text{H}_2\text{O}$) can easily be hydrated and converts to hexahydrite ($\text{MgSO}_4 \cdot 6\text{H}_2\text{O}$) and epsomite ($\text{MgSO}_4 \cdot 7\text{H}_2\text{O}$). When these minerals are desiccated, they do not convert back to kieserite,

but to an amorphous phase (Vaniman *et al.*, 2004). The presence of kieserite at certain locations on Mars (see below) indicates, therefore, that the climate probably never experienced obliquity-driven periods of hydration and desiccation in these regions (Vaniman *et al.*, 2004).

The presence of sulfates on Mars has been discussed for a long time. Theoretical considerations, partly based on terrestrial analog materials, suggested that Mars might host sulfates, and, for example, Burns (1987) predicted that jarosite might be the result of the interaction of groundwater with iron sulfides. Viking Lander measurements indicated a high sulfur content in the Martian soil, and Toulmin *et al.* (1977) proposed that sulfates might be the corresponding minerals. The SNC meteorites, which are thought to be samples of the Martian crust (e.g., Bogard and Johnson, 1983), contain sulfates (e.g., McSween, 1994), and spectroscopic observations also hinted toward sulfates (e.g., Pollack *et al.*, 1990; Blaney and McCord, 1995).

Major sulphate deposits have been found with OMEGA in the Valles Marineris and Terra Meridiani (Gendrin *et al.*, 2005), and the North Polar region (Langevin *et al.*, 2005a). In Valles Marineris, they are associated with layered deposits (**Figure 17**), and in the North Polar region they occur within a huge field of dark dunes. The MER confirmed the presence of sulfates and identified the Fe-sulfate jarosite, predicted to exist on Mars by

Burns (1987), in Meridiani Planum (Klingelhöfer *et al.*, 2004).

10.11.4.2.6 Carbonates

A warmer and denser greenhouse atmosphere on early Mars, as suggested by the geomorphic record of fluvial and lacustrine processes (see Section 10.11.3.1), would raise the question where the greenhouse gases have disappeared since then. A plausible way to remove CO₂ from the atmosphere would be the formation of carbonates, and the search for carbonates on Mars has been the subject of many investigations. However, despite the identification of carbonates in SNC meteorites (e.g., Bridges *et al.*, 2001), no carbonates have been found in Martian surface materials (Christensen *et al.*, 2001a; Bibring *et al.*, 2005).

10.11.4.2.7 Zeolites

Zeolites are common alteration products of basaltic ashes and palagonitic material on Earth, and they have been proposed as potential water-containing minerals on Mars. Spectral evidence for zeolites in the Martian dust has been presented by Ruff (2004), and possibly by Christensen *et al.* (2004b) in Mini-TES data taken in Gusev crater. Zeolites can easily be hydrated and dehydrated, and experimental studies have suggested that zeolites, if present on Mars, could undergo a diurnal cycle of hydration and dehydration with possible impact on the atmosphere (Bish *et al.*, 2003), but Tokano and Bish (2005) show that the diurnal atmospheric water cycle on Mars is unlikely to be affected by zeolites.



Figure 17 Layered deposit in Valles Marineris, Mars. The flank surfaces of this mountain were found to contain sulfates, based on spectra taken with the OMEGA instrument (Gendrin *et al.*, 2005). The sulfates seem to be associated with bright layers of unknown origin. The image is a false-color perspective view, based on HRSC stereo data (image width is about 70 km, and the height of the mountain is more than 4 km).

10.11.4.3 Geophysical Consequences

Based on experiments performed on dry peridotite systems water is capable of lowering the melting temperature of mantle material by as much as 500 K. The resulting melting curve for the water-saturated system has a negative *P-T* gradient at low pressures which becomes positive with increasing depth and therefore pressure. An amount as small as 0.1% of water would enable some partial melt to be formed, even if it would be impossible in the dry system. The amount of partial melting increases with increasing depth (Cattermole, 1989). Therefore, the effect of water on the solidus behavior can have a significant influence on the mantle of terrestrial planets. This is discussed in greater details in Chapter 10.09.

Small amounts of dissolved volatiles can have important effects on the density, viscosity, and crystallization of melts and magmas. Changing these properties influences the mechanical behavior of magmas, the differentiation, and the eruption style. The solubility of water in magmas depends on parameters like temperature, pressure, the presence of other volatiles (Williams and McBirney, 1979), and the presence of exchangeable cations, that is, those that are not in tetrahedral coordination sites (Burnham, 1979). A decrease in temperature and an increase in pressure both tend to increase the solubility of water in magma, and an increasing abundance of other volatiles decreases it (Cas and Wright, 1988). Consequently, the water contents of magmas vary widely (Table 1). In general, water is more soluble in highly differentiated magmas than in basaltic ones. For example, magmas from subduction zones are water rich, because sediments and secondary altered basalts provide water sources. It seems possible that a significant fraction of the water in magmas is added in crustal magma chambers (Kysor and O'Neil, 1984). It is, however, difficult to determine the original content of H₂O, in particular the amount of water injected into the stratosphere (Schmincke, 2000).

Mierdel *et al.* (2007) showed recently that the asthenosphere of Earth coincides with a zone where the water solubility in mantle minerals has a pronounced minimum. The minimum is due to a sharp decrease of water solubility in aluminous orthopyroxene with depth, whereas the water solubility in olivine continuously increases with pressure. Melting in the asthenosphere may therefore be related not to volatile enrichment but to a minimum in water solubility, which causes excess water to form a hydrous silicate melt.

Table 1 Approximate H₂O contents in magmas

Magma content	H ₂ O content (wt.%)
MORB ^a (tholeiites)	0.1–0.2
Island tholeiites	0.3–0.6
Alkali basalts	0.8–1.5
Subduction-zone basalts	2–3
Basanites and nephelinites	1.5–2
Andesites and dazites (island arcs)	1–2
Andesites and dazites (active continental margins)	2–4
Rhyolites	up to ~7

^a(From Schmincke (2000). After Fisher and Schmincke (1984) and Aoki *et al.* (1981)).
MORB, mid-ocean ridge basalts.

Dissolved water can change the density of and thereby influence, for example, the buoyancy. In an experiment with H₂O dissolved in a melt of albite composition (NaAlSi₃O₈), a partial molar volume of 14–20% cm³ mol⁻¹ has been measured for a temperature range from 750 to 950°C. Indirect estimates for the partial molar volume of water in rhyolitic or basaltic melts based on solubility data are generally at the lower end of the measurements for the albite melt (Hamilton and Oxtoby, 1986). Using the molecular weight of 18 g mol⁻¹ for H₂O the partial molar volume of ~14 cm³ mol⁻¹ the addition of dissolved water generates a component with a density of slightly greater than 1 g cm⁻³ compared to ~2.7 and ~2.3 g cm⁻³ for the anhydrous basaltic and rhyolitic melt, respectively (Wallace and Anderson, 2000). The addition of water decreases effectively the density of the melt. As an example adding 6% dissolved water to a rhyolitic melt decreases the density by about 5%.

The viscous behavior of melts is also discussed in Chapter 10.09. In general silicate melts exhibit Newtonian behavior, which means the rate of shear in the liquid is linearly proportional to the magnitude of applied shear stress. The viscosity increases rapidly with decreasing temperature. The viscosity is strongly dependent on the melt composition; with increasing SiO₂ content, the viscosity also increases. The difference in viscosity between a low-temperature rhyolitic melt to a high-temperature basaltic melt can be as much as eight orders of magnitude.

Adding dissolved water to a melt decreases the viscosity (e.g., Murase, 1962). The most likely cause is H₂O weakening or breaking apart (depolymerizing) the Si=O=Si bridges in the aluminosilicate framework of the melt (Wallace and Anderson, 2000). H₂O has a nonlinear effect on the viscosity as was shown in experiments (Hess and Dingwell, 1996). A proportion of 0–3% of water strongly reduces the viscosity. Adding 3 wt.% of water to a dry siliceous melt can decrease the viscosity by five orders of magnitude, while adding more than 3% of water has very little additional effect on the viscosity. For less than 3% H₂O, the water can dissolve in OH⁻ groups which bound to the aluminosilicate framework, thereby decreasing the viscosity very effectively. For higher amounts of water the saturation for OH⁻ groups is reached and additional water is dissolved as H₂O, having little effect on the aluminosilicate bonds; therefore, little additional change in the viscosity is observed.

Water is the most important volatile species controlling the acceleration and mass eruption rate of eruption columns. Pyroclastic explosive eruptions are mainly caused by the volumetric expansion of magmatic H₂O and the resulting fragmentation of magma. To illustrate this effect, we report an example given by Sparks *et al.* (1997): a volume of 1 m³ rhyolitic magma at 900°C and a H₂O content of 5 wt.% will expand to a 670 m³ of degassed mixture of water vapor and magma at 1 atm (1000 mbar) surface pressure. For small masses of up to 10–15 wt.% added surface water, the formation of an eruption column in an explosive event will be promoted (Koyaguchi and Woods, 1996). On the other hand, water contents that exceed 20–30 wt.% result in a dense, wet, and cold mixture, which requires a greater initial velocity to produce a buoyant column (Woods, 1995).

10.11.5 Evolution of Water and Climate

Of the terrestrial planets only Venus and Mars might have experienced in the past episodes in which a water-rich climate might have been possible.

Venus is at present an extremely dry and hot planet. The atmospheric water content is \sim 30 ppmv (e.g., see de Bergh *et al.* (1995) and the H₂O measurements tabulated in Fegley *et al.* (1997)) yielding a water partial pressure of \sim 3 mbar at the surface of Venus. For comparison Earth's average surface temperature is 288 K and its troposphere is much wetter with a water vapor content up to \sim 4% (see chapter 19 in Volume 1 of the Treatise on Geochemistry by Fegley). There are indications suggesting that Venus had more water in its past. This is based on the high deuterium/hydrogen (D/H) ratio of $(2.5 \pm 0.5) \times 10^{-2}$ in atmospheric water vapor reported by Donahue *et al.* (1997). It is 160 ± 32 times higher than the D/H ratio of 1.56 ± 10^{-4} in standard mean ocean water (SMOW) on Earth (Lodders and Fegley, 1998). One possible scenario to explain this ratio is enrichment by isotope fractionation. Hydrogen preferentially escapes through physical processes such as evaporation or diffusion, thus enriching the deuterium concentration (e.g., see Criss (1999)). The study of the D/H ratio by Donahue *et al.* (1982, 1997, and references therein) concluded that Venus once must have had the equivalent of at least a 4-m-deep global ocean and may have had as much as a 530-m-deep global ocean. This interpretation of the D/H ratio is complicated

by the fact that some interplanetary dust particles, meteorites, and comets also have a higher D/H ratio (Irvine *et al.*, 2000). Therefore, Venus might have accreted from material with a different D/H ratio than Earth. Further clues might be derived in the future by studying the surface mineralogy. While the extreme volcanic activity of Venus has most likely erased all morphological evidence for a water-rich past, there might be traces of hydrates minerals left as indications for a water-rich past (Johnson and Fegley, 2000).

If the D/H ratio is due to a water-rich past of the planet, Venus must have lost a large amount of water during its history. Rabette *et al.* (2002) proposed for earth a scenario in which an increase in sea surface temperatures above 27°C leads to strong evaporation loading the atmosphere with a critical amount of water vapor. If enough water vapor enters the troposphere a run-away greenhouse effect is initiated, causing even more water loss through evaporation from the ocean.

The ASPERA-4 instrument on Venus Express is studying the current oxygen and hydrogen escape rate from Venus which allows providing constraints on such escape models. Recent studies by Lammer *et al.* (2006) indicated that on Venus, due to its larger mass and size compared to Mars, the most relevant atmospheric escape processes of oxygen involve ions and are caused by the interaction with the solar wind. For the loss of hot H* atoms produced by photochemical reactions in the Venusian atmosphere Lammer (in press) found an average loss rate of about $3.8 \times 10^{25} \text{ s}^{-1}$. This is about two times lower than the estimated H⁺-ion outflow from Venus' nightside due to acceleration upward along magnetic field lines of ionospheric holes (Hartle and Grebowsky, 1993) in the order of about $7 \times 10^{25} \text{ s}^{-1}$. The loss rate for solar wind pick-up of H⁺ and O⁺ ions is about 1×10^{25} and $1.6 \times 10^{25} \text{ s}^{-1}$, respectively. Pick-up of H₂⁺ and He⁺ ions are negligible compared to the loss rates of atomic hydrogen and oxygen ions. Additional loss rates are found from detached plasma clouds which yield a loss of about $5 \times 10^{24} \text{ s}^{-1}$ to $1 \times 10^{25} \text{ s}^{-1}$ O⁺ ions, in agreement with the results of a two-dimensional global hybrid simulation of Terada *et al.* (2002) indicating that Venus loses O⁺ ions in the order of about $10^{25} \times \text{O}^+$ due to the Kelvin–Helmholtz instability. Heavy O atoms can only escape as neutrals by atmospheric sputtering in an order of about $6 \times 10^{24} \text{ s}^{-1}$ (Luhmann and Kozyra, 1991). The obtained results indicate that the ratio between H/O escape to space from the Venusian upper atmosphere is about 4, and

is in a much better agreement with the stoichiometrically H/O escape ratio of 2:1, which is not the case on Mars. The ongoing detailed analysis of the outflow of ions from the Venus upper atmosphere by the ASPERA-4 and VEX-MAG instruments aboard Venus Express will lead to more accurate atmospheric loss estimations and a better understanding of the planet's water inventory.

For Mars there are numerous studies on the atmospheric loss of water during the evolution of the planet. Chassefiere and Leblanc (2004) gave a summary of these studies and the resulting total loss of water from Mars. Zhang *et al.* (1993) and Luhmann *et al.* (1992) suggested an approach to evaluate the total amount of atmosphere lost to space along Mars' history (that is from 3.7 Gyr ago up to present time) based on knowledge of what should have been the early time of the Sun. The solar extreme-ultraviolet (XUV) flux was roughly three and six times higher than present XUV solar flux, 2.5 and 3.5 Gyr ago, respectively (Zahnle and Walker, 1982; Ayres, 1997). This results in a globally hotter Martian thermosphere 3.5 Gyr ago. Recent new model of the solar wind history (Wood *et al.*, 2002) suggested a 40 times more massive solar wind in the early solar system (3.5 Gyr ago) compared to the widely used model of Newkirk (1981). This has not been included in most models of the atmospheric loss so far. If this later process is neglected, the integrated loss to space associated with the estimates of Luhmann *et al.* (1992) for dissociative recombination and for pickup O⁺ corona, and with the estimates of Leblanc and Johnson (2001, 2002) for the sputtering, leads to a total loss to space of an H₂O ocean with a depth of 15 m covering the whole Mars surface. Lammer *et al.* (2003a, 2003b), using the same estimates, found a depth of 12 m. Previous estimates by Kass and Yung (1996) were of 50 m, and by Luhmann *et al.* (1992) and Luhmann (1997) of 30 m. Krasnopolsky and Feldman (2001) estimated the total loss of water to space based on D/H analysis and chemical model as being more than 30 m depth. All these values have to be compared to the most recent estimate of a 150-m-depth ocean of water covering Mars' surface during the Hesperian (from 3.5 Gyr ago) based on geomorphological analysis (see Section 10.11.3.4). Carr and Head (2003) concluded that a large majority of this water should have been trapped into the surface since only the equivalent of 20–30 m depth ocean is contained in the present polar caps.

A loss of volatiles from the Martian atmosphere and a corresponding climate change is consistent

with the evolution of the surface, which is recorded in the different ages of the various landforms. Fluvial and lacustrine landforms are common in ancient terrains (older than 3.5 Ga), whereas younger areas on Mars show only sparse evidence of a warmer climate that would allow the formation of such features. The morphologic evidence for a climate change billions of years ago is complementary to mineralogical evidence for a more water-rich early Mars. In addition to the atmospheric data presented in the above paragraph, the particular clues for this climate change have been presented in the sections on morphology and mineralogy. A detailed discussion of this major climate change, the nature and timing of which is still under debate, is beyond the scope of this paper, and the reader is referred to excellent studies by, for example, Pollack *et al.* (1987), Kasting (1991), Pollack (1991), Fanale *et al.* (1992), Jakosky and Jones (1997), Jakosky and Phillips (2001), and Bogard *et al.* (2001). In the following paragraph, we briefly summarize the discussion on possible, much more recent climate changes.

For Mars already early studies on Mariner 9 and Viking images showed glacial and permafrost landforms. The higher-resolution data of recent missions like MGS, Mars Odyssey, and Mars Express provided a surprising wealth of data showing details of many cold-climate phenomena. Many of them appear to be young and display a latitude-dependent, nonrandom distribution on a global scale. Such a distribution is typical for climatically controlled surface features. Although young, many of these landforms require the existence of ice and/or water at latitudes where it is not physically stable today. Moreover, morphological structures like the mid-latitude cover deposit seem to be in a state of degradation. These observations suggest that recent climate excursions triggered precipitation in mid-latitudes and perhaps even in low latitudes, probably as snow. One plausible reason for such climate changes in recent time could be the variable obliquity of Mars' rotational axis (Ward, 1973, 1974; Carr, 1982; Touma and Wisdom, 1993; Laskar and Robutel, 1993; Jakosky *et al.*, 1995; Laskar *et al.*, 2004; Armstrong *et al.*, 2004). In contrast to Earth, whose obliquity is stabilized by the Moon (Laskar *et al.*, 1993), the obliquity of Mars is characterized by chaotic behavior and has a high probability to reach maximum values of 60–80° and minimum values of ~0° (Laskar *et al.*, 2004). At periods of high obliquity, the mean annual insolation at the polar regions is enhanced, and more volatiles tend to sublimate.

Therefore, the polar regions experience a net loss of water, which is transported equatorwards and precipitated as snow in mid- and low-latitudes (Levrard *et al.*, 2004). This snow could have mantled the surface, producing a ‘smoothed’ appearance of the topography. Where local topographic conditions favored enhanced precipitation, larger amounts of snow could have accumulated and formed glaciers (Forget *et al.*, 2006). In periods of low obliquity, the mean insolation at mid- and low-latitudes is enhanced and the ice-related landforms are degraded.

10.11.6 Summary and Outlook

Water is a highly vital, flexible, changeable, and moveable substance that alters and evolves other materials. The H₂O molecule is a major component of planets, moons, asteroids, and comets and acts in atmospheres, on surfaces, and in the subsurface as a weathering agent and transportation and sedimentation medium. It can also influence the rheology of materials and thereby influences the mantle transport processes (*see* Chapter 10.09).

Water is the main ingredient for life. As a polar molecule it is an ideal solvent and supports several important processes for sustaining life. As water is crucial for habitable environments, its physical conditions constrain habitable zones (*see* Chapter 10.12).

Remote-sensing data show abundant evidence of fluvial and lacustrine processes which have been active on early Mars on a global scale (Noachian) with a declining rate of activity and processes limited to regional and local scales (Hesperian). These landforms also imply a large past reservoir of water on the surface and a past climate on Mars which allowed liquid water to be stable on the surface for longer timescales. Today Mars is a cold and dry planet, but glacial, periglacial, and fluvial landforms indicate a geologically recent occurrence of water ice in the near-surface regolith, on the surface, and at least episodically in the liquid phase on the surface. The formation of these young landforms on Mars might be due to recent climate changes caused by the variable obliquity of Mars’ rotational axis.

Strictly even on Mars all evidence for water is indirect, mainly based on remote-sensing observation of either the O–H bond (Christensen *et al.*, 2001a; Formisano *et al.*, 2005; Bibring *et al.*, 2005) or of the hydrogen atom (Feldman *et al.*, 2005). The MER added morphological and indirect geochemical

evidence, based on water-related alteration processes. Only in the polar areas we have direct observations of spectral features of water ice (Bibring *et al.*, 2005; Hansen *et al.*, 2005). Radar systems as, currently, MARSIS on Mars Express (Picardi *et al.*, 2005) and SHARAD on MRO allow for the first time to study the subsurface structure and identify potential water or ice deposits below the surface. In combination with the data from the neutron and GRS this allows a more accurate global inventory of water on Mars today. The advances in remote sensing will allow a much lower detection limit and a higher spatial resolution for mapping of water-related features. The HiRISE and GRISM instruments on MRO will allow studying the morphological and mineralogical traces of water activity at a very high spatial resolution. Thereby it will be possible to trace the history of water on Mars with a much higher confidence. Even given all these advancements up to now no *in situ* instrument has actually ‘touched’ a water molecule on Mars. To further the understanding of the current-day water cycle on Mars it is necessary to probe the hypothesized or remotely sensed water reservoirs directly. This includes direct measurements of the soil humidity at different landing sites, lander in the polar areas probing the ice and directly analyzing its composition, and access to the subsurface to probe for possible subsurface ice deposits. Highly mobile platforms as the NASA rover Mars Science Lab (MSL) or the ESA rover ExoMars, equipped with drilling systems, neutron spectrometer, and GPR will open a new dimension in the analysis of present and past water on Mars.

Assessing the proposed ice deposits close to the poles of Mercury will be possible with the instrument suites of the NASA mission MESSENGER and the ESA mission BepiColombo (McNutt *et al.*, 2004; Solomon *et al.*, 2001). Both spacecrafts are equipped with gamma-ray and neutron spectrometer which will allow studying the spatial distribution and abundance of hydrogen in the regolith. Furthermore, spectrometer on both spacecrafts covering the wavelengths range from UV to thermal infrared will perform a systematic search for ice or alternative explanations for the radar bright material. In combination with significant improvements in the topography the synergetic use of data provided by both spacecrafts will allow to make a comprehensive argument for or against water ice in the shadowed regions at the poles of Mercury.

For the Moon the detection of water ice in permanently shadowed craters close to the poles is less certain. At the same time for the future (human)

exploration of the Moon these potential deposits are of great importance. Therefore, this will be a prime scientific priority for the new exploration program of the Moon. Already the first mission, the NASA Lunar Reconnaissance Orbiter, will carry a neutron spectrometer and in addition a small subsatellite (Lunar Crater Observation and Sensing Satellite – LCROSS) studying the deposits in Shackleton crater by creating an artificial impact. Further mission plans include landing and sample returns from these areas.

References

- Aharonson O, Zuber MT, Neumann GA, and Head JW (1998) Mars: Northern hemisphere slopes and slope distributions. *Geophysical Research Letters* 25: 4413–4416.
- Albee AL, Arvidson RE, Palluconi F, and Thorpe T (2001) Overview of the Mars Global Surveyor mission. *Journal of Geophysical Research* 106: 23291–23316.
- Albee AL, Palluconi FD, and Arvidson RE (1998) Mars Global Surveyor mission: Overview and status. *Science* 279: 1671–1672.
- Allen CC (1979) Areal distribution of Martian rampart craters. *Icarus* 39: 111–123.
- Andreas EL (2007) New estimates for the sublimation rate for ice on the moon. *Icarus* 186(1): 24–30.
- Ansan V and Mangold N (2006) New observations of Warrego Valles, Mars: Evidence for precipitation and surface runoff. *Planetary and Space Science* 54: 219–242.
- Aoki K, Ishiwaka K, and Kanisawa S (1981) Fluorine geochemistry of basaltic rocks from continental and oceanic regions and petrogenetic application. *Contributions to Mineralogy and Petrology* 76: 53–59.
- Armstrong JC, Levy CB, and Quinn T (2004) A 1 Gyr climate model for Mars: New orbital statistics and the importance of seasonally resolved polar processes. *Icarus* 171: 255–271.
- Arnold JR (1979) Ice in the lunar polar regions. *Journal of Geophysical Research* 84: 5659–5668.
- Ayres TR (1997) Evolution of the solar ionizing flux. *Journal of Geophysical Research* 102: 1641–1651.
- Baker VR (1982) *The Channels of Mars*, 198pp. Austin, TX: University of Texas Press.
- Baker VR (2001) Water and the Martian landscape. *Nature* 412: 228–236.
- Baker VR, Benito G, and Rudy AN (1993) Paleohydrology of late Pleistocene superflooding, Altay Mountains, Siberia. *Science* 259: 348–350.
- Baker VR, Carr MH, Gulick VC, Williams CR, and Marley MS (1992) Channels and valley networks. In: Kieffer HH, Jakosky BM, Snyder CW, and Matthews MS (eds.) *Mars*, pp. 493–522. Tucson, AZ: University of Arizona Press.
- Baker VR and Kochel RC (1979) Martian channel morphology – Maja and Kasei Valles. *Journal of Geophysical Research* 84: 7961–7983.
- Baker VR and Milton DJ (1974) Erosion by catastrophic floods on Mars and Earth. *Icarus* 23: 27–41.
- Baker VR and Nummedal D (1978) *NASA Planetary Geology Program: The Channeled Scabland*. Washington, DC: NASA.
- Baker VR, Strom RG, Gulick VC, Kargel JS, and Komatsu G (1991) Ancient oceans, ice sheets and the hydrological cycle on Mars. *Nature* 352: 589–594.
- Ballou EV, Wood PC, Wydeven T, Lehwall ME, and Mack R (1978) Chemical interpretation of Viking Lander 1 life detection experiment. *Nature* 271: 644–645.
- Balme M, Mangold N, Baratoux D, et al. (2006) Orientation and distribution of recent gullies in the southern hemisphere of Mars: Observations from high resolution stereo camera/Mars Express (HRSC/MEX) and Mars Orbiter camera/Mars Global Surveyor (MOC/MGS) data. *Journal of Geophysical Research* 111: E05001 (doi:10.1029/2005JE002607).
- Bandfield JL, Glotch TD, and Christensen PR (2003) Spectroscopic identification of carbonate minerals in the Martian dust. *Science* 301: 1084–1087.
- Banerdt WB and Vidal A (2001) Surface drainage on Mars. *Lunar and Planetary Science* XXXII: 1488 (CD-ROM).
- Barlow NG (2004) Martian subsurface volatile concentrations as a function of time: Clues from layered ejecta craters. *Geophysical Research Letters* 31: L05703 (doi:10.1029/2003GL019075).
- Barlow NG (2005) A review of Martian impact crater ejecta structures and their implications for target properties. In: Kenkmann T, Hötz F, and Deutsch A (eds.) *Geological Society of America Special Paper*, Vol. 384: *Large Meteorite Impacts III*, pp. 433–442. Boulder, CO: GSA Publication.
- Barlow NG, Boyce JM, Costard FM, et al. (2000) Standardizing the nomenclature of Martian impact crater ejecta morphologies. *Journal of Geophysical Research* 105: 26733–26738.
- Barlow NG and Bradley TL (1990) Martian impact craters – Correlations of ejecta and interior morphologies with diameter, latitude, and terrain. *Icarus* 87: 156–179.
- Barlow NG, Koroshetz J, and Dohm JM (2001) Variations in the onset diameter for Martian layered ejecta morphologies and their implications for subsurface volatile reservoirs. *Geophysical Research Letters* 28: 3095–3098.
- Barnouin-Jha OS and Schultz PH (1998) Lobateness of impact ejecta deposits from atmospheric interactions. *Journal of Geophysical Research* 103: 25739–25756.
- Barsch D (1996) *Rock Glaciers*, 331pp. Berlin, Germany: Springer.
- Bart GD (2007) Comparison of small lunar landslides and Martian gullies. *Icarus* 187: 417–421.
- Bell JF (1996) Iron, sulfate, carbonate, and hydrated minerals on Mars. In: Burns RG, Dyar MD, McCammon C, and Schaefer MW (eds.) *Geochemical Society Special Publication*, Vol. 5: *Mineral Spectroscopy: A Tribute of Roger G. Burns*, pp. 359–380. London: The Geochemical Society.
- Bell JF, Squyres SW, Herkenhoff KE, et al. (2003) Mars exploration rover athena panoramic camera (Pancam) investigation. *Journal of Geophysical Research* 108: 8063 (doi:10.1029/2003JE002070).
- Benison KC and Bowen BB (2006) Acid saline lake systems give clues about past environments and the search for life on Mars. *Icarus* 183: 225–229.
- Benn DI and Evans DJA (1998) *Glaciers and Glaciation*, 734pp. London: Arnold.
- Bennett MR and Glasser NF (1996) *Glacial Geology: Ice Sheets and Landforms*, 364pp. Chichester, UK: Wiley.
- Berman DC, Hartmann WK, and Crown DA (2003) Debris aprons, channels, and volcanoes in the Reull Vallis region of Mars. *Lunar and Planetary Science* XXXIV: 1879.
- Bertini HW (1969) Intranuclear-cascade calculation of the secondary nucleon spectra from nucleon-nucleus interactions in the energy range 340 to 2900 MeV and comparisons with experiment. *Physical Review* 188: 1711–1730.
- Bhattacharya JP, Payenberg THD, Lang SC, and Bourke M (2005) Dynamic river channels suggest a long-lived Noachian crater lake on Mars. *Geophysical Research Letters* 32: L10201 (doi:10.1029/2005GL022747).

- Bibring J-P and Erard S (2001) The Martian surface composition. *Space Science Reviews* 96: 293–316.
- Bibring J-P, Langevin Y, Gendrin A, et al. (2005) Mars surface diversity as revealed by the OMEGA/Mars Express observations. *Science* 307: 1576–1581.
- Bibring J-P, Langevin Y, Mustard JF, et al. (2006) Global mineralogical and aqueous Mars history derived from OMEGA/Mars Express data. *Science* 312: 400–404.
- Bibring J, Langevin Y, Poulet F, et al. (2004a) Perennial water ice identified in the south polar cap of Mars. *Nature* 428: 627–630.
- Bibring J-P, Langevin Y, Soufflot A, Combes C, and Cara C (1989) Results from the ISM experiment. *Nature* 341: 591–593.
- Bibring J-P, Soufflot A, Berthé M, et al. (2004b) OMEGA: Observatoire pour la minéralogie, l'eau, les glaces et l'activité. *European Space Agency Special Publication* 1240: 37–49.
- Bish DL, William Carey J, Vaniman DT, and Chipera SJ (2003) Stability of hydrous minerals on the Martian surface. *Icarus* 164: 96–103.
- Bishop JL (2005) Hydrated minerals on Mars. In: Tokano T (ed.) *Water on Mars and Life*, pp. 65–96. Berlin, Germany: Springer.
- Bishop JL, Murad E, Lane MD, and Mancinelli RL (2004) Multiple techniques for mineral identification on Mars: A study of hydrothermal rocks as potential analogues for astrobiology sites on Mars. *Icarus* 169: 311–323.
- Bishop JL, Pieters CM, Burns RG, Edwards JO, Mancinelli RL, and Froeschl H (1995) Reflectance spectroscopy of ferric sulfate-bearing montmorillonites as Mars soil analog materials. *Icarus* 117: 101–119.
- Blaney DL and McCord TB (1995) Indications of sulphate minerals in the Martian soil from Earth-based spectroscopy. *Journal of Geophysical Research* 100: 14433–14442.
- Bloom AL (1991) *Geomorphology*, 2nd edn., 532pp. Englewood Cliffs, NY: Prentice Hall.
- Bodnar RJ (2001) PTX Phase equilibria in the H_2O-CO_2 -salt system at Mars near-surface conditions. *Lunar and Planetary Science XXXII*: 1689 (CD-ROM).
- Bogard DD, Clayton RN, Marti K, Owen T, and Turner G (2001) Martian volatiles: Isotopic composition, origin, and evolution. *Space Science Reviews* 96: 425–458.
- Bogard DD and Johnson P (1983) Martian gases in an antarctic meteorite. *Science* 221: 651–654.
- Boyce JM, Mouginis-Mark P, and Garbeil H (2005) Ancient oceans in the northern lowlands of Mars: Evidence from impact crater depth/diameter relationships. *Journal of Geophysical Research* 110: 3008 (doi:10.1029/2004JE002328).
- Boyce JM and Witbeck NE (1980) Distribution of thermal gradient values in the equatorial region of Mars based on impact crater morphology. *Reports of Planetary Geology Program, NASA TM-82385*, pp. 140–143.
- Boynton WV, Feldman WC, Squyres SW, et al. (2002) Distribution of hydrogen in the near surface of Mars: Evidence for subsurface ice deposits. *Science* 297: 81–85.
- Brakenridge GR (1993) Modern shelf ice, equatorial Aeolis quadrangle, Mars. *Lunar and Planetary Institute Conference*, pp. 175–176.
- Brass GW (1980) Stability of brines on Mars. *Icarus* 42: 20–28.
- Bretz H (1923) The channeled scabland of the Columbia plateau. *Journal of Geology* 31: 617–649.
- Bridges JC, Catling DC, Saxton JM, Swindle TD, Lyon IC, and Grady MM (2001) Alteration assemblages in Martian meteorites: Implications for near-surface processes. *Space Science Reviews* 96: 365–392.
- Bridges NT and Lackner CN (2006) Northern hemisphere Martian gullies and mantled terrain: Implications for near-surface water migration in Mars' recent past. *Journal of Geophysical Research* 111: 9014 (doi:10.1029/2006JE002702).
- Brunauer S, Emmett PH, and Teller E (1938) Adsorption of gases in multimolecular layers. *Journal of American Chemical Society* 60: 309–319.
- Bruno BC, Fagents SA, Hamilton CW, Burr DM, and Baloga SM (2006) Identification of volcanic rootless cones, ice mounds, and impact craters on Earth and Mars: Using spatial distribution as a remote sensing tool. *Journal of Geophysical Research* 111: E06017 (doi:10.1029/2005JE002510).
- Bruno BC, Fagents SA, Thordarson T, Baloga SM, and Pilger E (2004) Clustering within rootless cone groups on Iceland and Mars: Effect of nonrandom processes. *Journal of Geophysical Research* 109: E07009 (doi:10.1029/2004JE002273).
- Bun Tseung J-MW and Soare RJ (2006) Thermokarst and related landforms in western Utopia Planitia, Mars: Implications for near-surface excess ice. *Lunar and Planetary Science XXXVII*: 1414 (CD-ROM).
- Burnham CW (1979) The importance of volatile constituents. In: Yoder HS (ed.) *The Evolution of Igneous Rocks*, pp. 439–482. Princeton, NJ: Princeton University Press.
- Burns RG (1970) *Mineralogical Applications of Crystal Field Theory*, 224pp. New York: Cambridge University Press.
- Burns RG (1987) Ferric sulfates on Mars. *Journal of Geophysical Research* 92: 570–574.
- Burr DM (2003) Hydraulic modelling of Athabasca Valles, Mars. *Hydrological Sciences Journal* 48: 655–664.
- Burr DM, Grier JA, McEwen AS, and Keszthelyi LP (2002a) Repeated aqueous flooding from the Cerberus Fossae: Evidence for very recently extant, deep groundwater on Mars. *Icarus* 159: 53–73.
- Burr DM, McEwen AS, and Sakimoto SEH (2002b) Recent aqueous floods from the Cerberus Fossae, Mars. *Geophysical Research Letters* 29: 1013 (doi:10.1029/2001GL013345).
- Burr DM, Soare RJ, Bun Tseung J-MW, and Emery JP (2005) Young (late Amazonian), near-surface, ground ice features near the equator, Athabasca Valles, Mars. *Icarus* 178: 56–73.
- Bussey DBJ, Spudis PD, and Robinson MS (1999) Illumination conditions at the lunar south pole. *Geophysical Research Letters* 26: 1187–1190.
- Butler BJ (1997) The migration of volatiles on the surface of Mercury and the Moon. *Journal of Geophysical Research* 102: 19283–19292.
- Butler BJ, Muhselman DO, and Slade MA (1993) Mercury – Full-disk radar images and the detection and stability of ice at the North Pole. *Journal of Geophysical Research* 98: 15003–15023.
- Cabrol NA and Grin EA (1999) Distribution, classification, and ages of Martian impact crater lakes. *Icarus* 142: 160–172.
- Cabrol NA and Grin EA (2001) The evolution of lacustrine environments on Mars: Is Mars only hydrologically dormant? *Icarus* 149: 291–328.
- Cabrol NA, Grin EA, and Dawidowicz G (1996) Ma'adim Vallis revisited through new topographic data: Evidence for an ancient intravalley lake. *Icarus* 123: 269–283.
- Cabrol NA, Grin EA, and Dawidowicz G (1997) A model of outflow generation by hydrothermal underpressure drainage in volcano-tectonic environment, Shalbatana Vallis (Mars). *Icarus* 125: 455–464.
- Cabrol NA, Grin EA, Landheim R, Kuzmin RO, and Greeley R (1998) Duration of the Ma'adim Vallis/Gusev crater hydrogeologic system. *Mars* 133: 98–108.
- Cabrol NA, Grin EA, and Pollard WH (2000) Possible frost mounds in an ancient Martian lake bed. *Icarus* 145: 91–107.

- Calvin WM (1997) Variation of the 3- μm absorption feature on Mars: Observations over eastern Valles Marineris by the Mariner 6 infrared spectrometer. *Journal of Geophysical Research* 102: 9097–9108.
- Campbell DB, Campbell BA, Carter LM, Margot J-L, and Stacy NJS (2006) No evidence for thick deposits of ice at the lunar south pole. *Nature* 443: 835–837.
- Campbell DB, Chandler JF, Ostro SJ, Pettengill GH, and Shapiro II (1978) Galilean satellites – 1976 radar results. *Icarus* 34: 254–267.
- Carr MH (1974) The role of lava erosion in the formation of lunar rilles and Martian channels. *Icarus* 22: 1–23.
- Carr MH (1979) Formation of Martian flood features by release of water from confined aquifers. *Journal of Geophysical Research* 84: 2995–3007.
- Carr MH (1982) Periodic climate change on Mars: Review of evidence and effects on distribution of volatiles. *Icarus* 50: 129–139.
- Carr MH (1983) Stability of streams and lakes on Mars. *Icarus* 56: 476–495.
- Carr MH (1995) The Martain drainage system and the origin of valley networks and fretted channels. *Journal of Geophysical Research* 100: 7479–7507.
- Carr MH (1996) *Water on Mars*, 229pp. New York: Oxford University Press.
- Carr MH (2001) Mars Global Surveyor observations of Martian fretted terrain. *Journal of Geophysical Research* 106: 23571–23594.
- Carr MH (2002) Elevations of water-worn features on Mars: Implications for circulation of groundwater. *Journal of Geophysical Research* 107: 5131 (doi:10.1029/2002JE001845).
- Carr MH and Chuang FC (1997) Martian drainage densities. *Journal of Geophysical Research* 102: 9145–9152.
- Carr MH and Clow GD (1981) Martian channels and valleys – Their characteristics, distribution, and age. *Icarus* 48: 91–117.
- Carr MH, Crumpler LS, Cutts JA, Greeley R, Guest JE, and Masursky H (1977) Martian impact craters and emplacement of ejecta by surface flow. *Journal of Geophysical Research* 82: 4055–4065.
- Carr MH and Head JW (2003) Oceans on Mars: An assessment of the observational evidence and possible fate. *Journal of Geophysical Research* 108: 5042 (doi:10.1029/2002JE001963).
- Carr MH and Schaber GG (1977) Martian permafrost features. *Journal of Geophysical Research* 82: 4039–4054.
- Cas RAF and Wright JV *Volcanic Successions*, 528pp. London: Unwin Hyman.
- Catling DC (1999) A chemical model for evaporites on early Mars: Possible sedimentary tracers of the early climate and implications for exploration. *Journal of Geophysical Research* 104: 16453–16470.
- Catling DC and Moore JM (2003) The nature of coarse-grained crystalline hematite and its implications for the early environment of Mars. *Icarus* 165: 277–300.
- Cattermole PJ (1989) *Planetary Volcanism – A study of Volcanic Activity in the Solar System*, 443pp. Chichester, UK: Ellis Horwood.
- Chassefiere E and Leblanc F (2004) Mars atmospheric escape and evolution; interaction with the solar wind. *Planetary and Space Science* 52(11): 1039–1058.
- Christiansen EH (1989) Lahars in the Elysium region of Mars. *Geology* 17: 203–206.
- Christensen PR (2003) Formation of recent Martian gullies through melting of extensive water-rich snow deposits. *Nature* 422: 45–48.
- Christensen PR, Anderson DL, Chase SC, et al. (1992) Thermal emission spectrometer experiment – Mars observer mission. *Journal of Geophysical Research* 97: 7719–7734.
- Christensen PR, Bandfield JL, Clark RN, et al. (2000) Detection of crystalline hematite mineralization on Mars by the thermal emission spectrometer: Evidence for nearsurface water. *Journal of Geophysical Research* 105: 9623–9642.
- Christensen PR, Bandfield JL, Hamilton VE, et al. (2001a) Mars Global Surveyor thermal emission spectrometer experiment: Investigation description and surface science results. *Journal of Geophysical Research* 106: 23823–23872.
- Christensen PR, Mehall GL, Silverman SH, et al. (2003) Miniature thermal emission spectrometer for the Mars exploration rovers. *Journal of Geophysical Research (Planets)* 108: 8064 (doi:10.1029/2003JE002117).
- Christensen PR, Morris RV, Lane MD, Bandfield JL, and Malin MC (2001b) Global mapping of Martian hematite mineral deposits: Remnants of water-driven processes on early Mars. *Journal of Geophysical Research* 106: 23873–23886.
- Christensen PR, Ruff SW, Ferguson RL, et al. (2004b) Initial results from the Mini-TES experiment in Gusev crater from the spirit rover. *Science* 305: 837–842 (doi:10.1126/science.1100564).
- Christensen PR, Wyatt MB, Glotch TD, et al. (2004a) Mineralogy at Meridiani Planum from the Mini-TES experiment on the Opportunity rover. *Science* 306: 1733–1739.
- Christensen PR, Wyatt MB, Glotch TD, et al. (2004c) Mineralogy at Meridiani Planum from the Mini-TES experiment on the Opportunity rover. *Science* 306: 1733–1739.
- Chuang FC and Crown DA (2005) Surface characteristics and degradational history of debris aprons in the Tempe Terra/ Mareotis fossae region of Mars. *Icarus* 179: 24–42.
- Clark BC, III, Castro AJ, Rowe CD, et al. (1977) The Viking X ray fluorescence experiment – Analytical methods and early results. *Journal of Geophysical Research* 82: 4577–4594.
- Clark RN (1981) Water frost and ice – The near-infrared spectral reflectance 0.65–2.5 microns. *Journal of Geophysical Research* 86: 3087–3096.
- Clark RN (1999) Spectroscopy of rocks and minerals, and principles of spectroscopy. In: Rencz AN (ed.) *Manual of Remote Sensing*, Vol. 3: *Remote Sensing for the Earth Sciences*, pp. 3–58. New York: John Wiley and Sons.
- Clark RN, Swayze GA, Singer RB, and Pollack JB (1990) High-resolution reflectance spectra of Mars in the 2.3-micron region – Evidence for the mineral scapolite. *Journal of Geophysical Research* 95: 14463–14480.
- Clifford SM (1987) Polar basal melting on Mars. *Journal of Geophysical Research* 92: 9135–9152.
- Clifford SM (1993) A model for the hydrologic and climatic behavior of water on Mars. *Journal of Geophysical Research* 98: 10973–11016.
- Clifford SM and Parker TJ (2001) The evolution of the Martian hydrosphere: Implications for the fate of a primordial ocean and the current state of the northern plains. *Icarus* 154: 40–79.
- Clow GD (1987) Generation of liquid water on Mars through the melting of a dusty snowpack. *Icarus* 72: 95–127.
- Colaprete A and Jakosky BM (1998) Ice flow and rock glaciers on Mars. *Journal of Geophysical Research* 103: 5897–5910.
- Coleman NM (2003) Aqueous flows carved the outflow channels on Mars. *Journal of Geophysical Research* 108: 5039 (doi:10.1029/2002JE001940).
- Coleman NM (2004) Ravi Vallis, Mars paleoflood origin and genesis of secondary chaos zones. *Lunar and Planetary Science XXXV: 1299 (CD-ROM)*.
- Coleman NM (2005) Martian megaflood-triggered chaos formation, revealing groundwater depth, cryosphere thickness, and crustal heat flux. *Journal of Geophysical Research* 110: E12S20 (doi:10.1029/2005JE002419).
- Costa JE (1984) Physical geomorphology of debris flows. In: Costa JE and Fleisher PJ (eds.) *Developments and Applications of Geomorphology*, pp. 268–317. Berlin, Germany: Springer.

- Costard F, Aguirre-Puente J, Greeley R, and Makhoul N (1999) Martian fluvial-thermal erosion: Laboratory simulation. *Journal of Geophysical Research* 104(E6): 14091–14098 (doi: 10.1029/1999JE900020).
- Costard F, Forget F, Mangold N, and Peulvast JP (2002) Formation of recent Martian debris flows by melting of near-surface ground ice at high obliquity. *Science* 295: 110–113.
- Costard FM (1989) The spatial distribution of volatiles in the Martian hydrolithosphere. *Earth, Moon and Planets* 45: 265–290.
- Costard FM and Baker VR (2001) Thermokarst landforms and processes in Ares Vallis, Mars. *Geomorphology* 37: 289–301.
- Costard FM and Kargel JS (1995) Outwash plains and thermokarst on Mars. *Icarus* 114: 93–112.
- Craddock RA and Howard AD (2002) The case for rainfall on a warm, wet early Mars. *Journal of Geophysical Research* 107: E5111 (doi:10.1029/2001JE001505).
- Criss RE (1999) *Principles of Stable Isotope Distribution*, 254pp. New York: Oxford University Press.
- Cutts JA (1973) Nature and Origin of Martian Polar Layered Deposits. *Bulletin of the American Astronomical Society* 78: 4231–4249.
- Cutts JA and Blasius KR (1981) Origin of Martian outflow channels – The eolian hypothesis. *Journal of Geophysical Research* 86: 5075–5102.
- de Bergh C, Bezard B, Crisp D, et al. (1995) Water in the deep atmosphere of Venus from high-resolution spectra of the night side. *Advances in Space Research* 15: 79–88.
- De Hon R (1992) Martian lake basins and lacustrine plains. *Earth, Moon, and Planets* 56: 95–122.
- Di Achille G, Marinangeli L, Ori GG, et al. (2006a) Geological evolution of the Tyrras Vallis paleolacustrine system, Mars. *Journal of Geophysical Research* 111: E04003 (doi:10.1029/2005JE002561).
- Di Achille G, Ori GG, Reiss D, et al. (2006b) A steep fan at Coprates Catena, Valles Marineris, Mars, as seen by HRSC data. *Geophysical Research Letters* 33: L07204 (doi:10.1029/2005GL025435).
- Di Achille G, Ori GG, and Reiss D (in press) Evidence for Late Hesperian lacustrine activity in Shalbatana Vallis, Mars. *Journal of Geophysical Research*.
- Dickson JL, Head JW, and Kreslavsky M (in press) Martian gullies in the southern mid-latitudes of Mars: Evidence for climate-controlled formation of young fluvial features based upon local and global topography. *Icarus*.
- Donahue TM, Grinspoon DH, Hartle RE, and Hodges RR (1997) Ion/neutral escape of hydrogen and deuterium: Evolution of water. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and Solar Wind Environment*, pp. 385–414. Tucson, AZ: University of Arizona Press.
- Donahue TM, Hoffman JH, Hodges RR, and Watson AJ (1982) Venus was wet – A measurement of the ratio of deuterium to hydrogen. *Science* 216: 630–633.
- Drake DM, Feldman WC, and Jakosky BM (1988) Martian neutron leakage spectra. *Journal of Geophysical Research* 93: 6353–6368.
- Easterbrook DJ (1999) *Surface Processes and Landforms*, 2nd edn., 546pp. New York: Prentice Hall.
- Edgett KS, Malin MC, Williams RME, and Davis SD (2003) Polar- and middle-latitude Martian gullies: A view from MGS MOC after 2 Mars years in the mapping orbit. *Lunar and Planetary Science* XXXIV: 1038 (CD-ROM).
- Edgett KS and Parker TJ (1997) Water on early Mars: Possible subaqueous sedimentary deposits covering ancient cratered terrain in western Arabia and Sinus Meridiani. *Geophysical Research Letters* 24: 2897–2900.
- Embleton C and King CAM (1975) *Glacial Geomorphology*, 573pp. London: Arnold.
- Estermann I (1955) Gases at low densities. In: Rossini FD (ed.) *Thermodynamics and Physics of Matter*, pp. 736–776. Princeton, NJ: Princeton University Press.
- Evans DJA (2003) *Glacial Landsystems*, 532pp. London: Arnold.
- Fairén AG, Fernández-Remolar D, Dohm JM, Baker VR, and Amils R (2004) Inhibition of carbonate synthesis in acidic oceans on early Mars. *Nature* 431: 423–426.
- Fanale FP (1999) Mars: Atmosphere and volatile history. In: Weissman P, McFadden L, and Johnson T (eds.) *Encyclopedia of the Solar System*, pp. 277–290. Orlando, FL: Academic Press.
- Fanale FP and Cannon WA (1971) Adsorption on the Martian regolith. *Nature* 230: 502–504.
- Fanale FP and Cannon WA (1974) Exchange of absorbed H₂O and CO₂ between the regolith and atmosphere of Mars caused by changes in surface insolation. *Journal of Geophysical Research* 79: 3397–3402.
- Fanale FP, Postawko SE, Pollack JB, Carr MH, and Pepin RO (1992) Mars – Epochal climate change and volatile history. In: Kieffer HH, Jakosky BM, Synder CW, and Matthews MS (eds.) *Mars*, pp. 1135–1179. Tucson, AZ: University of Arizona Press.
- Farmer CB, Davies DW, and Laporte DD (1976) Mars – Northern summer ice cap – Water vapor observations from Viking 2. *Science* 194: 1339–1341.
- Farmer CB and Doms PE (1979) Global seasonal variation of water vapor on Mars and the implications of permafrost. *Journal of Geophysical Research* 84(13): 2881–2888.
- Farmer VC (1974) *The Infra-Red Spectra of Minerals*, 539pp. London: Mineralogical Society.
- Fassett CI and Head JW, III (2006) Valleys on Hecates Tholus, Mars: Origin by basal melting of summit snowpack. *Planetary and Space Science* 54: 370–378.
- Fegley B, Jr, Klingelhöfer G, Lodders K, and Widemann T (1997) Geochemistry of surface atmosphere interactions on Venus. In: Bouger SW, Hunten DM, and Philips RJ (eds.) *Venus II: Geology, Geophysics, Atmosphere, and SolarWind Environment*, pp. 591–636. Tucson, AZ: University of Arizona Press.
- Feldman WC, Boynton WV, Jakosky BM, and Mellon MT (1993) Redistribution of subsurface neutrons caused by ground ice on Mars. *Journal of Geophysical Research* 98: 20855–20870.
- Feldman WC, Boynton WV, Tokar RL, et al. (2002) Global distribution of neutrons from Mars: Results from Mars Odyssey. *Science* 297: 75–78.
- Feldman WC and Drake DM (1986) A Doppler filter technique to measure the hydrogen content of planetary surfaces. *Nuclear Instruments and Methods in Physics Research* 245: 182–190.
- Feldman WC, Lawrence DJ, Elphic RC, et al. (2000) Polar hydrogen deposits on the Moon. *Journal of Geophysical Research* 105: 4175–4196.
- Feldman WC, Maurice S, Binder AB, Barracough BL, Elphic RC, and Lawrence DJ (1998) Fluxes of fast and epithermal neutrons from *Lunar Prospector*: Evidence for water ice at the lunar poles. *Science* 281: 1496.
- Feldman WC, Prettyman TH, Maurice S, et al. (2004) Global distribution of near-surface hydrogen on Mars. *Journal of Geophysical Research* 109: E09006 (doi:10.1029/2003JE002160).
- Feldman WC, Prettyman TH, Maurice S, et al. (2005) Topographic control of hydrogen deposits at low latitudes to midlatitudes of Mars. *Journal of Geophysical Research* 110: E11009 (doi:10.1029/2005JE002452).
- Fisher RV and Schmincke H-U (1984) *Pyroclastic Rocks*, 472pp. Berlin, Germany: Springer.

- Flasar FM and Goody RM (1976) Diurnal behaviour of water on Mars. *Planetary and Space Science* 24: 161–181.
- Fletcher NH (1970) *The Chemical Physics of Ice*, 271pp. Cambridge, UK: Cambridge University Press.
- Forget F, Haberle RM, Montmessin F, Levrard B, and Head JW (2006) Formation of glaciers on Mars by atmospheric precipitation at high obliquity. *Science* 311: 368–371.
- Formisano V, Angrilli F, Arnold G, et al. (2005) The planetary fourier spectrometer (PFS) onboard the European Mars Express mission. *Planetary and Space Science* 53: 963–974.
- Forsythe RD and Zimbelman JR (1995) A case for ancient evaporite basins on Mars. *Journal of Geophysical Research* 100: 5553–5563.
- French HM (1996) *The Periglacial Environment*, 2nd edn., 341pp. Essex, UK: Addison Wesley Longman.
- Gaidos EJ (2001) Note: Cryovolcanism and the recent flow of liquid water on Mars. *Icarus* 153: 218–223.
- Gendrin A, Mangold N, Bibring J, et al. (2005) Sulfates in Martian layered terrains: The OMEGA/Mars Express view. *Science* 307: 1587–1591.
- Ghatan GJ and Head JW (2002) Candidate subglacial volcanoes in the south polar region of Mars: Morphology, morphometry, and eruption conditions. *Journal of Geophysical Research* 107: 5048 (doi:10.1029/2001JE001519).
- Ghatan GJ, Head JW, and Wilson L (2005) Mangala Valles, Mars: Assessment of early stages of flooding and downstream flood evolution. *Earth, Moon and Planets* 96: 1–57 (doi:10.1007/s11038-005-9009-y).
- Ghatan GJ and Zimbelman JR (2006) Paucity of candidate coastal constructional landforms along proposed shorelines on Mars: Implications for a northern lowlandsfilling ocean. *Icarus* 185(1): 171–196.
- Gilmore MS and Phillips EL (2002) Role of aquiclude in formation of Martian gullies. *Geology* 30: 1107–1110.
- Glen JW (1955) The creep of polycrystalline ice. *Proceedings of the Royal Society of London, Series A, Mathematical, Physical and Engineering Sciences* 228(1175): 519–538.
- Goetz W, Bertelsen P, Binou CS, et al. (2005) Indication of drier periods on Mars from the chemistry and mineralogy of atmospheric dust. *Nature* 436: 62–65.
- Goldspiel JM and Squyres SW (2000) Groundwater sapping and valley formation on Mars. *Icarus* 148: 176–192.
- Goldspiel JM, Squyres SW, and Jankowski DG (1993) Topography of small Martian valleys. *Icarus* 105: 479–500.
- Golombek MP and Bridges NT (2000) Erosion rates on Mars and implications for climate change: Constraints from the pathfinder landing site. *Journal of Geophysical Research* 105: 1841–1854.
- Golombek MP, Grant JA, Crumpler LS, et al. (2006) Erosion rates at the Mars exploration rover landing sites and long-term climate change on Mars. *Journal of Geophysical Research* 111: E12S10 (doi:10.1029/2006JE002754).
- Golombek MP, Grant JA, Parker TJ, et al. (2003) Selection of the Mars exploration rover landing sites. *Journal of Geophysical Research* 108: 8072 (doi:10.1029/2003JE002074).
- Grant JA (2000) Valley formation in Margaritifer Sinus, Mars, by precipitation-recharged ground-water sapping. *Geology* 28: 223–226.
- Greeley R (1985) *Planetary Landscapes*, 286pp. London: Allen & Unwin.
- Greeley R and Fagents SA (2001) Icelandic pseudocraters as analogs to some volcanic cones on Mars. *Journal of Geophysical Research* 106: 20527–20546.
- Greeley R and Guest JE (1987) *US Geological Survey Miscellaneous Investigation Series, I-1802-B: Geologic Map of the Eastern Equatorial Region of Mars*. Flagstaff, AZ: US Geological Survey.
- Greeley R and Spudis PD (1981) Volcanism on Mars. *Reviews of Geophysics and Space Physics* 19: 13–41.
- Grimm RE (2002) Low-frequency electromagnetic exploration for groundwater on Mars. *Journal of Geophysical Research* 107: 5006 (doi:10.1029/2001JE001504).
- Grin EA and Cabrol NA (1997) Limnologic analysis of Gusev crater paleolake, Mars. *Icarus* 130: 461–474.
- Grosswald MG (1998) New approach to the Ice Age paleohydrology of northern Eurasia. In: Benito G, Baker VR, and Gregory KJ (eds.) *Paleohydrology and Environmental Change*, pp. 199–214. New York: John Wiley & Sons.
- Grotzinger JP, Arvidson RE, Bell JF, et al. (2005) Stratigraphy and sedimentology of a dry to wet eolian depositional system, Burns formation, Meridiani Planum. *Mars, Earth and Planetary Science Letters* 240: 11–72.
- Gulick VC (2001) Origin of the valley networks on Mars: A hydrological perspective. *Geomorphology* 37: 241–268.
- Gulick VC and Baker VR (1989) Fluvial valleys and Martian palaeoclimates. *Nature* 341: 514–516.
- Gulick VC and Baker VR (1990) Origin and evolution of valleys on Martian volcanoes. *Journal of Geophysical Research* 95: 14325–14344.
- Haberle RM, McKay CP, Schaeffer J, et al. (2001) On the possibility of liquid water on present-day Mars. *Journal of Geophysical Research* 106: 23317–23326.
- Hambrey MJ (1994) *Glacial Environments*, 296pp. London: UCL Press.
- Hamilton DL and Oxtoby S (1986) Solubility of water in albite-melt determined by the weight-loss method. *Journal of Geology* 94: 626–630.
- Hamilton WC and Ibers JA (1968) *Hydrogen Bonding in Solids; Methods of Molecular Structure Determination*. New York: W. A. Benjamin.
- Hansen G, Giuranna M, Formisano V, et al. (2005) PFS-MEX observation of ices in the residual south polar cap of Mars. *Planetary and Space Science* 53: 1089–1095.
- Hapke B and Blewett D (1991) Coherent backscatter model for the unusual radar reflectivity of icy satellites. *Nature* 352: 46–47.
- Hargraves RB, Collinson DW, Arvidson RE, and Spitzer CR (1977) The Viking magnetic properties experiment – Primary mission results. *Journal of Geophysical Research* 82: 4547–4558.
- Harmon JK, Perillat PJ, and Slade MA (2001) High-resolution radar imaging of Mercury's north pole. *Icarus* 149: 1–15.
- Harmon JK and Slade MA (1992) Radar mapping of Mercury – Full-disk images and polar Anomalies. *Science* 258: 640–643.
- Harrison KP and Grimm RE (2004) Tharsis recharge: A source of groundwater for Martian outflow channels. *Geophysical Research Letters* 31: 14703 (doi:10.1029/2004GL020502).
- Harrison KP and Grimm RE (2005) Groundwater-controlled valley networks and the decline of surface runoff on early Mars. *Journal of Geophysical Research* 110: E12S16 (doi:10.1029/2005JE002455).
- Hartle RE and Grebowsky JM (1993) Light ion flow in the nightside ionosphere of Venus. *Journal of Geophysical Research* 98: 7437–7445.
- Hartmann WK (2001) Martian seeps and their relation to youthful geothermal activity. *Space Science Reviews* 96: 405–410.
- Hasegawa S, Murakawa K, Ishiguro M, et al. (2003) Evidence of hydrated and/or hydroxylated minerals on the surface of asteroid 4 Vesta. *Geophysical Research Letters* 30: 2123 (doi:10.1029/2003GL018627).
- Hasted JB (1973) *Aqueous Dielectrics*, 302pp. London: Chapman and Hall.
- Hauber E, van Gasselt S, Chapman MG, and Neukum G (2007) Geomorphic evidence for former lobate debris aprons at low

- latitudes on Mars: Indicators of the Martian paleoclimate. *Lunar and Planetary Science XXXVIII*: 1666.
- Hauber E, van Gasselt S, Ivanov B, et al. (2005) Discovery of a flank caldera and very young glacial activity at Hecates Tholus, Mars. *Nature* 434: 356–361.
- Head JW, Kreslavsky M, Hiesinger H, et al. (1998) Oceans in the past history of Mars: Tests for their presence using Mars Orbiter laser altimeter (MOLA) data. *Geophysical Research Letters* 25: 4401–4404.
- Head JW, Kreslavsky MA, and Pratt S (2002) Northern lowlands of Mars: Evidence for widespread volcanic flooding and tectonic deformation in the Hesperian Period. *Journal of Geophysical Research* 107: 5003 (doi:10.1029/2000JE001445).
- Head JW and Marchant DR (2003) Cold-based mountain glaciers on Mars: Western Arisia Mons. *Geology* 31: 641–644.
- Head JW, Marchant DR, Agnew MC, Fassett CI, and Kreslavsky MA (2006) Extensive valley glacier deposits in the northern mid-latitudes of Mars: Evidence for Late Amazonian obliquity-driven climate change. *Earth and Planetary Science Letters* 241: 663–671.
- Head JW, Mustard JF, Kreslavsky MA, Milliken RE, and Marchant DR (2003a) Recent ice ages on Mars. *Nature* 426: 797–802.
- Head JW, Neukum G, Jaumann R, et al. (2005) Tropical to midlatitude snow and ice accumulation, flow and glaciation on Mars. *Nature* 434: 346–351.
- Head JW and Roth R (1976) Mars pedestal crater escarpments: Evidence for ejecta-related emplacement. *IPI Contribution No. 259: Papers Presented to the Symposium on Planetary Cratering Mechanics*, pp. 50–52. Houston, TX: Lunar Science Institute.
- Head JW, Wilson L, and Mitchell KL (2003b) Generation of recent massive water floods at Cerberus Fossae, Mars by dike emplacement, cryospheric cracking, and confined aquifer groundwater release. *Geophysical Research Letters* 30: 1577 (doi:10.1029/2003GL017135).
- Head JW, III, Kreslavsky M, Hiesinger H, and Pratt S (1999) Northern seas and oceans in the past history of Mars: New evidence from Mars Orbiter laser altimeter (MOLA) data. *Lunar and Planetary Science XXX*: 1352 (CD-ROM).
- Heggy E, Paillou P, Ruffie G, Malezieux JM, Costard F, and Grandjean G (2001) On water detection in the Martian subsurface using sounding radar. *Icarus* 154: 244–257.
- Helbert J, Reiss D, Hauber E, and Benkhoff J (2005) Limits on the burial depth of glacial ice deposits on the flanks of Hecates Tholus, Mars. *Geophysical Research Letters* 32: L17201 (doi:10.1029/2005GL023712).
- Heldmann JL and Mellon MT (2004) Observations of Martian gullies and constraints on potential formation mechanisms. *Icarus* 168: 285–304.
- Heldmann JL, Toon OB, Pollard WH, et al. (2005) Formation of Martian gullies by the action of liquid water flowing under current Martian environmental conditions. *Journal of Geophysical Research* 110: 5004 (doi:10.1029/2004JE002261).
- Herkenhoff KE and Murray BC (1990) High-resolution topography and albedo of the South Polar layered deposits on Mars. *Journal of Geophysical Research* 95: 14511–14529.
- Hess KU and Dingwell DB (1996) Viscosities of hydrous leucogranitic melts: A non- Arrhenian model. *American Mineralogist* 81: 1297–1300.
- Hiesinger H and Head JW (2000) Characteristics and origin of polygonal terrain in southern Utopia Planitia, Mars: Results from Mars Orbiter laser altimeter and Mars Orbiter camera data. *Journal of Geophysical Research* 105(E5): 11999–12022 (doi:10.1029/1999JE001193).
- Hoffman N (2000a) White Mars: A new model for Mars' surface and atmosphere based on CO₂. *Icarus* 146: 326–342.
- Hoffman N (2000b) Ideas about the surface runoff features on Mars. *Science* 290: 711–714.
- Hoffman N (2002) Active polar gullies on Mars and the role of carbon dioxide. *Astrobiology* 2: 313–323 (doi:10.1089/153110702762027899).
- Howard AD and Kochel RC (1988) Introduction to cuesta landforms and sapping processes on the Colorado Plateau. In: Howard AD, Kochel RC, and Holt HR (eds.) *NASA Special Publication, Vol. 491: Sapping Features of the Colorado Plateau*, pp. 6–56. Washington, DC: NASA.
- Howard AD, Moore JM, and Irwin RP (2005) An intense terminal epoch of widespread fluvial activity on early Mars. 1: Valley network incision and associated deposits. *Journal of Geophysical Research* 110: E12S14 (doi:10.1029/2005JE002459).
- Huggett RJ (2003) *Fundamentals of Geomorphology*, 386pp. London: Routledge.
- Hynek BM and Phillips RJ (2003) New data reveal mature, integrated drainage systems on Mars indicative of past precipitation. *Lunar and Planetary Science XXXIV*: 1842 (CDROM).
- Ingersoll AP, Svitek T, and Murray BC (1992) Stability of polar frosts in spherical bowlshaped craters on the moon, Mercury, and Mars. *Icarus* 100: 40–47.
- Irvine WM, Schloerb FP, Crovisier J, Fegley B, Jr, and Mumma MJ (2000) Comets: A link between interstellar and nebular chemistry. In: Boss A, Manning V, and Russell S (eds.) *Protostars and Planets IV*, p. 1159. Tucson, AZ: University of Arizona Press.
- Irwin RP and Howard AD (2002) Drainage basin evolution in Noachian Terra Cimmeria, Mars. *Journal of Geophysical Research* 107: 5056 (doi:10.1029/2001JE001818).
- Irwin RP, Howard AD, Craddock RA, and Moore JM (2005a) An intense terminal epoch of widespread fluvial activity on early Mars. 2: Increased runoff and paleolake development. *Journal of Geophysical Research* 110: E12S12 (doi:10.1029/2005JE002460).
- Irwin RP, III, Craddock RA, and Howard AD (2005b) Interior channels in Martian valley networks: Discharge and runoff production. *Geology* 33: 489–492.
- Irwin RP, III, Craddock RA, Howard AD, and Maxwell TA (2004) Channels in Martian valley networks: Discharge and runoff production. *Second Conference on Early Mars: Geologic, Hydrologic, and Climatic Evolution and the Implications for Life*, 8040. Lunar and Planetary Science (CDROM).
- Ivanov MA and Head JW (2001) Chryse Planitia, Mars: Topographic configuration, outflow channel continuity and sequence, and tests for hypothesized ancient bodies of water using Mars Orbiter laser altimeter (MOLA) data. *Journal of Geophysical Research* 106: 3275–3296.
- Jahn A (1975) *Problems of the Periglacial Zone*, 223pp. Warschau, UK: Polish Scientific Publications.
- Jakosky BM (1985) The seasonal cycle of water on Mars. *Space Science Reviews* 41: 131–200.
- Jakosky BM, Henderson BG, and Mellon MT (1995) Chaotic obliquity and the nature of the Martian climate. *Journal of Geophysical Research* 100: 1579–1584.
- Jakosky BM and Jones JH (1997) The history of Martian volatiles. *Reviews of Geophysics* 35: 1–16.
- Jakosky BM and Phillips RJ (2001) Mars' volatile and climate history. *Nature* 412: 237–244.
- James PB, Kieffer HH, and Paige DA (1992) The seasonal cycle of carbon dioxide on Mars. In: Kieffer HH, Jakosky BM, Synder CW, and Matthews MS (eds.) *Mars*, pp. 934–968. Tucson, AZ: University of Arizona Press.
- Jänicke J, Bish DL, Möhlmann DTF, and Stach H (2006) Investigation of the water sorption properties of

- Mars-relevant micro- and mesoporous minerals. *Icarus* 180: 353–358.
- Jankowski DG and Squyres SW (1992) The topography of impact craters in ‘softened’ terrain on Mars. *Icarus* 100: 26–39 (doi:10.1016/0019-1035(92)90015-Y).
- Jankowski DG and Squyres SW (1993) Softened impact craters on Mars – Implications for ground ice and the structure of the Martian megaregolith. *Icarus* 106: 365–379 (doi:10.1006/icar.1993.1178).
- Jaumann R, Reiss D, Frei S, et al. (2005) Interior channels in Martian valleys: Constraints on fluvial erosion by measurements of the Mars Express high resolution stereo camera. *Geophysical Research Letters* 32: L16203 (doi:10.1029/2005GL023415).
- Jerolmack DJ, Mohrig D, Zuber MT, and Byrne S (2004) A minimum time for the formation of Holden Northeast fan, Mars. *Geophysical Research Letters* 31: L21701 (doi:10.1029/2004GL021326).
- Johnson CL, Solomon SC, Head JW, Phillips RJ, Smith DE, and Zuber MT (2000) Lithospheric loading by the northern polar cap on Mars. *Icarus* 144: 313–328.
- Johnson NM and Fegley B (2000) Water on Venus: New insights from tremolite decomposition. *Icarus* 146: 301–306.
- Jöns H-P (1984) Sedimentary basins and mud flows in the northern lowlands of Mars. *Lunar and Planetary Institute Conference*, pp. 417–418.
- Jöns H-P (1985) Late sedimentation and late sediments in the northern lowlands on Mars. *Lunar and Planetary Institute Conference*, pp. 414–415.
- Jöns H-P (1986) Arcuate ground undulations, gelifluxionlike features, and ‘Front Tori’ in the northern lowlands of Mars – What do they indicate? *Lunar and Planetary Institute Conference*, pp. 404–405.
- Jöns H-P (2002) Junge/Rezente nichtäolische exogene dynamik auf dem Mars: Großflächige Beeinflussung durch zerfallende CO₂-hydrate? *Zeitschrift Fur Geologische Wissenschaften* 6: 403–421.
- Joung D, Poulet F, Mustard JF, et al. (2006) Observation of 3 m hydration feature on Mars from OMEGA-MEx data. *Lunar and Planetary Science XXXVII*: 1741 (CD-ROM).
- Kargel JS, Baker VR, Beget JE, et al. (1995) Evidence of ancient continental glaciation in the Martian northern plains. *Journal of Geophysical Research* 100: 5351–5368.
- Kargel JS and Strom RG (1992) Ancient glaciation on Mars. *Geology* 20: 3–7.
- Kass DM and Yung YL (1996) The loss of atmosphere from Mars response. *Science* 274: 1932–1933.
- Kasting JF (1991) CO₂ condensation and the climate of early Mars. *Icarus* 94: 1–13.
- Kasting JF, Whitmire DP, and Reynolds RT (1993) Habitable zones around main sequence stars. *Icarus* 101: 108–128.
- Kerr R (2006) At last, Methane lakes on Saturn’s icy moon Titan – But no seas. *Science* 313: 598.
- Keszthelyi L, Thordarson T, McEwen A, et al. (2004) Icelandic analogs to Martian flood lavas. *Geochemistry, Geophysics, Geosystems* 5: Q11014.
- Kieffer HH (1979) Mars south polar spring and summer temperatures – A residual CO₂ frost. *Journal of Geophysical Research* 84: 8263–8288.
- Kieffer HH, Martin TZ, Chase SC, Miner ED, and Palluconi FD (1976) Martian north pole summer temperatures – Dirty water ice. *Science* 194: 1341–1344.
- Kieffer HH, Titus TN, Mullins KF, and Christensen PR (2000) Mars south polar spring and summer behavior observed by TES: Seasonal cap evolution controlled by frost grain size. *Journal of Geophysical Research* 105: 9653–9700.
- Kleinhanz MG (2005) Flow discharge and sediment transport models for estimating a minimum timescale of hydrological activity and channel and delta formation on Mars. *Journal of Geophysical Research* 110: 12003 (doi:10.1029/2005JE002521).
- Klingelhöfer G, Morris RV, Bernhardt B, et al. (2003) Athena MIMOS II Mössbauer spectrometer investigation. *Journal of Geophysical Research* 108: 8067 (doi:10.1029/2003JE002138).
- Klingelhöfer G, Morris RV, Bernhardt B, et al. (2004) Jarosite and hematite at Meridiani Planum from Opportunity’s Mössbauer spectrometer. *Science* 306: 1740–1745.
- Kneissl T (2006) *Verbreitung und Azimutabhängigkeiten von Erosionsrinnen in der nördlichen Hemisphäre des Mars*, 81 pp. Unpublished Master Thesis, Martin-Luther-Universität Halle, Wittenberg.
- Komar PD (1979) Comparisons of the hydraulics of water flows in Martian outflow channels with flows of similar scale on earth. *Icarus* 37: 156–181.
- Komar PD (1980) Modes of sediment transport in channelized water flows with ramifications to the erosion of the Martian outflow channels. *Icarus* 42: 317–329.
- Koyaguchi T and Woods AW (1996) On the formation of eruption columns following explosive mixing of magma and surface-water. *Journal of Geophysical Research* 101: 5561–5574.
- Krasnopolsky VA and Feldman PD (2001) Detection of molecular hydrogen in the atmosphere of Mars. *Science* 294: 1914–1917.
- Kreslavsky MA and Head JW (1999) Kilometer-scale slopes on Mars and their correlation with geologic units: Initial results from Mars Orbiter laser altimeter (MOLA) data. *Journal of Geophysical Research* 104: 21911–21924.
- Kreslavsky MA and Head JW (2000) Kilometer-scale roughness of Mars: Results from MOLA data analysis. *Journal of Geophysical Research* 105: 26695–26712.
- Kreslavsky MA and Head JW (2002) Fate of outflow channel effluents in the northern lowlands of Mars: The Vastitas Borealis Formation as a sublimation residue from frozen ponded bodies of water. *Journal of Geophysical Research (Planets)* 107: 5121 (doi:10.1029/2001JE001831).
- Kuzmin RO (1980) Determination of frozen soil depth on Mars from the morphology of fresh craters. *Akademii Nauk SSSR Doklady* 252: 1445–1448.
- Kuzmin RO (2005) Ground ice in the Martian regolith. In: Tokano T (ed.) *Advances in Astrobiology and Biogeophysics: Water on Mars and Life*, pp. 155–189. Berlin, Germany: Springer.
- Kuzmin RO, Bobina NN, Zabalueva EV, and Shashkina VP (1988) Structural inhomogeneities of the Martian cryolithosphere. *Solar System Research* 22: 121–133.
- Kuzmin RO, Bobina NN, Zabalueva EV, and Shashkina VP (1989) Structural irregularities of the upper layers of the Martian cryolithosphere. In: *Space Chemistry and Comparative Planetology*, pp. 80–95. Dordrecht, The Netherlands: Kluwer.
- Kuzmin RO and Zabalueva EV (1998) On salt solutions in the Martian cryolithosphere. *Astronomicheskii Vestnik* 32: 187.
- Kuzmin RO and Zabalueva EV (2003) Polygonal terrains on Mars: Preliminary results of global mapping of their spatial distribution. *Lunar and Planetary Science XXXIV*: 1912 (CD-ROM).
- Kyser TK and O’Neil JR (1984) Hydrogen isotope systematics of submarine basalts. *Geochimica et Cosmochimica Acta* 48: 2123–2133.
- Lachenbruch AH (1962) Mechanics of thermal contraction cracks and ice-wedge polygons in permafrost. *Geological Society of America, Special Paper* 70: 69.
- Laity JE and Malin MC (1985) Sapping processes and the development of theatre-headed valley networks on the Colorado Plateau. *Geological Society of America, Bulletin* 96: 203–217.

- Lajeunesse E, Quantin C, Allemand P, and Delacourt C (2006) New insights on the runoff of large landslides in the Valles Marineris canyons, Mars. *Geophysical Research Letters* 33(4): L04403 (doi:10.1029/2005GL025168).
- Lamb MP, Howard AD, Johnson J, Whipple KX, Dietrich WE, and Perron JT (2006) Can springs cut canyons into rock? *Journal of Geophysical Research* 111: 7002 (doi:10.1029/2005JE002663).
- Lammer H, Kolb C, Penz T, Amerstorfer UV, Biernat HK, and Bodiselitsch B (2003b) Estimation of the past and present Martian water-ice reservoirs by isotopic constraints on exchange between the atmosphere and the surface. *International Journal of Astrobiology* 2: 1–8.
- Lammer H, Lichtenegger HIM, Biernat HK, et al. (2006) Loss of hydrogen and oxygen from the upper atmosphere of Venus. *Planetary and Space Science* 54(13–14): 1445–1456.
- Lammer H, Lichtenegger HIM, Kolb C, et al. (2003a) Loss of water from Mars implications for the oxidation of the soil. *Icarus* 165: 9–25.
- Langaman PD, McEwen AS, Keszthelyi LP, and Thordarson T (2001) Rootless cones on Mars indicating the presence of shallow equatorial ground ice in recent times. *Geophysical Research Letters* 28: 2365–2368.
- Langevin Y, Poulet F, Bibring J-P, and Gondet B (2005a) Sulfates in the north polar region of Mars detected by OMEGA/Mars Express. *Science* 307: 1584–1586.
- Langevin Y, Poulet F, Bibring J-P, Schmitt B, Douté S, and Gondet B (2005b) Summer evolution of the north polar cap of Mars as observed by OMEGA/Mars Express. *Science* 307: 1581–1584.
- Langmuir L (1915) Chemical reactions at low pressures. *Journal of American Chemical Society* 37: 1139–1167.
- Lanz J (2004) Geometrische, morphologische und stratigraphische Untersuchungen ausgewählter Outflow Channel der Circum-Chryse-Region, Mars, mit Methoden der Fernerkundung, 180pp. Dissertation, Free University Berlin, DLRForschungsbericht 2004-02.
- Laskar J, Correia ACM, Gastineau M, Joutel F, Levrard B, and Robutel P (2004) Long term evolution and chaotic diffusion of the insolation quantities of Mars. *Icarus* 170: 343–364.
- Laskar J, Joutel F, and Robutel P (1993) Stabilization of the earth's obliquity by the moon. *Nature* 361: 615–617.
- Laskar J and Robutel P (1993) The chaotic obliquity of the planets. *Nature* 361: 608–612.
- Leask HJ, Wilson L, and Mitchell KL (2006a) Formation of Ravi Vallis outflow channel, Mars: Morphological development, water discharge, and duration estimates. *Journal of Geophysical Research* 111: 8070 (doi:10.1029/2005JE002550).
- Leask HJ, Wilson L, and Mitchell KL (2006b) Formation of Aromatum chaos, Mars: Morphological development as a result of volcano–ice interactions. *Journal of Geophysical Research* 111: 8071 (doi:10.1029/2005JE002549).
- Leblanc F and Johnson RE (2001) Sputtering of the Martian atmosphere by solar wind pick-up ions. *Planetary and Space Science* 49: 645–656.
- Leblanc F and Johnson RE (2002) Role of molecules in pick-up ion sputtering of the Martian atmosphere. *Journal of Geophysical Research* 107: 5010 (doi:10.1029/2000JE001473).
- Lee P, Cockell CS, Marinova MM, McKay CP, and Rice JW (2001) Snow and ice melt flow features on Devon Island, Nunavut, Arctic Canada as possible analogs for recent slope flow features on Mars. *Lunar and Planetary Science XXXII*: 1809 (CD-ROM).
- Levrard B, Forget F, Montmessin F, and Laskar J (2004) Recent ice-rich deposits formed at high latitudes on Mars by sublimation of unstable equatorial ice during low obliquity. *Nature* 431: 1072–1075.
- Leverington DW (2004) Volcanic rilles, streamlined islands, and the origin of outflow channels on Mars. *Journal of Geophysical Research* 109: 10011 (doi:10.1029/2004JE002311).
- Li H, Robinson MS, and Jurdy DM (2005) Origin of Martian northern hemisphere midlatitude lobate debris aprons. *Icarus* 176: 382–394.
- Lodders and Fegley (1998) *Treatise on geochemistry* volume 1, Chapter 19.
- Lozinski W (1909) Über die mechanische Verwitterung der Sandsteine im gemässigten Klima. *Bulletin de l'Academie des Sciences de Gacouie class des Sciences Mathématique et Naturelles* 1: 1–25.
- Lucchitta BK (1979) Landslides in Valles Marineris, Mars. *Journal of Geophysical Research* 84: 8097–8113.
- Lucchitta BK (1981) Mars and earth – Comparison of coldclimate features. *Icarus* 45: 264–303.
- Lucchitta BK (1982) Ice sculpture in the Martian outflow channels. *Journal of Geophysical Research* 87: 9951–9973.
- Lucchitta BK (1984) Ice and debris in the fretted terrain, Mars. *Journal of Geophysical Research Supplement* 89: 409–418.
- Lucchitta BK (1985) Geomorphologic evidence for ground ice on Mars. In: Klinger J, Benest D, Dollfus A, and Smoluchowski A (eds.) *Ices in the Solar System*, pp. 583–604. Dordrecht, The Netherlands: D. Reidel Publishing.
- Lucchitta BK (1987) Valles Marineris, Mars – Wet debris flows and ground ice. *Icarus* 72: 411–429.
- Lucchitta BK (2001) Antarctic ice streams and outflow channels on Mars. *Geophysical Research Letters* 28: 403–406 (doi:10.1029/2000GL011924).
- Lucchitta BK, Ferguson HM, and Summers C (1986) Sedimentary deposits in the northern lowland plains, Mars. *Journal of Geophysical Research* 91: 166–174.
- Lucchitta BK, Isbell NK, and Howington-Kraus A (1994) Topography of Valles Marineris: Implications for erosional and structural history. *Journal of Geophysical Research* 99: 3783–3798.
- Lucey P, Korotev RL, Gillis JJ, et al. (2006) New views of the moon. In: Jolliff BL, Wieczorek MA, Shearer CK, and Neal CR (eds.) *Reviews in Mineralogy and Geochemistry, Vol. 60: Understanding the Lunar Surface and Space–Moon Interactions*, pp. 83–219. Corpus Christi, TX: World Scientific.
- Luhmann JG (1997) Correction to 'The ancient oxygen exosphere of Mars implications for atmospheric evolution'. *Journal of Geophysical Research* 102: 1637.
- Luhmann JG, Johnson RE, and Zhang MHG (1992) Evolutionary impact of sputtering of the Martian atmosphere by O⁺ pick-up ions. *Geophysical Research Letters* 19: 2151–2154.
- Luhmann JG and Kozyra JU (1991) Dayside pickup oxygen ion precipitation at Venus and Mars: Spatial distributions, energy deposition and consequences. *Journal of Geophysical Research* 96: 5457–5467.
- Mackay JR (1973) The growth of pingos, western Arctic coast, Canada. *Canadian Journal of Earth Sciences* 10(6): 979–1004.
- Mackay JR (1979) Pingos of the Tuktoyaktuk Peninsula area, Northwest Territories. *Geographie Physique et Quaternaire* 33: 3–61.
- Malin MC and Carr MH (1999) Groundwater formation of Martian valleys. *Nature* 397: 589–591.
- Malin MC and Edgett KS (1999) Oceans or seas in the Martian northern lowlands: High-resolution imaging tests of proposed coastlines. *Geophysical Research Letters* 26: 3049–3052.
- Malin MC and Edgett KS (2000c) Evidence for recent groundwater seepage and surface runoff on Mars. *Science* 288: 2330–2335.

- Malin MC and Edgett KS (2000b) Sedimentary rocks of early Mars. *Science* 290: 1927–1937.
- Malin MC and Edgett KS (2000c) Constraints on overland fluid transport through Martian valley networks. *Lunar and Planetary Science XXXVI*: 1189 (CD-ROM).
- Malin MC and Edgett KS (2001) Mars Global Surveyor Mars Orbiter camera: Interplanetary cruise through primary mission. *Journal of Geophysical Research* 106: 23429–23570.
- Malin MC and Edgett KS (2003) Evidence for persistent flow and aqueous sedimentation on early Mars. *Science* 302: 1931–1934.
- Malin MC, Edgett KS, Posiolova LV, McColley SM, and Noe Dobrea EZ (2006) Present-day impact cratering rate and contemporary gully activity on Mars. *Science* 314: 1573–1577.
- Manga M (2004) Martian floods at Cerberus Fossae can be produced by groundwater discharge. *Geophysical Research Letters* 31: 2702 (doi:10.1029/2003GL018958).
- Mangold N (2003) Geomorphic analysis of lobate debris aprons on Mars at Mars Orbiter camera scale: Evidence for ice sublimation initiated by fractures. *Journal of Geophysical Research* 108(E4): 8021 (doi: 10.1029/2002JE001885).
- Mangold N (2005) High latitude patterned grounds on Mars: Classification, distribution and climatic control. *Icarus* 174: 336–359.
- Mangold N and Ansan V (2006) Detailed study of an hydrological system of valleys, a delta and lakes in the Southwest Thaumasia region, Mars. *Icarus* 180: 75–87.
- Mangold N, Costard F, and Forget F (2003) Debris flows over sand dunes on Mars: Evidence for liquid water. *Journal of Geophysical Research* 108: 5027 (doi:10.1029/2002JE001958).
- Mangold N, Maurice S, Feldman WC, Costard F, and Forget F (2004) Spatial relationships between patterned ground and ground ice detected by the neutron spectrometer on Mars. *Journal of Geophysical Research* 109(E8): E08001 (doi:10.1029/2004JE002235).
- Mangold N, Quantin C, Ansan V, Delacourt C, and Allemand P (2004) Evidence for precipitation on Mars from dendritic valleys in the Valles Marineris Area. *Science* 305: 78–81.
- Manning R (1891) On the flow of water in open channels and pipes. *Transactions of the Institute of Civil Engineers, Ireland* 20: 161–207.
- Margot JL, Campbell DB, Jurgens RH, and Slade MA (1999) Topography of the lunar poles from radar interferometry: A survey of cold trap locations. *Science* 284: 1658–1660.
- Mars Channel Working Group (1983) Channels and valleys on Mars. *Geological Society of America, Bulletin* 94: 1035–1054.
- Masursky H (1973) An overview of geological results from Mariner 9. *Journal of Geophysical Research* 78: 4009–4030.
- Masursky H, Boyce JM, Dial AL, Schaber GG, and Strobell ME (1977) Classification and time of formation of Martian channels based on Viking data. *Journal of Geophysical Research* 82: 4016–4038.
- Masursky H, Chapman MG, Dial AL, Jr, and Strobell ME (1986) Episodic channeling by volcanic flows in Mangala Valles region, Mars. pp.459–461 *Reports of Planetary Geology Program, NASA TM-88383*.
- Max MD and Clifford SM (2001) Initiation of Martian outflow channels: Related to the dissociation of gas hydrate? *Geophysical Research Letters* 28: 1787–1790.
- McCauley JF (1978) US Geological Survey Miscellaneous Investigation Series. Map I-897: *Geological map of the coprates quadrangle of Mars*. Flagstaff, AZ: US Geological Survey.
- McCord TB and Adams JB (1969) Spectral reflectivity of Mars. *Science* 163: 1058–1060.
- McEwen AS (1989) Mobility of large rock avalanches: Evidence from Valles Marineris. *Geology* 17: 1111–1114.
- McEwen AS, Eliason EM, Bergstrom JW, et al. (in press) MRO's high resolution imaging science experiment (HiRISE). *Journal of Geophysical Research*.
- McGill GE (1985) Age and origin of large Martian polygons. *Lunar and Planetary Institute Conference*, pp. 534–535.
- McGill GE (1986) The giant polygons of Utopia, northern Martian plains. *Geophysical Research Letters* 13: 705–708.
- McNutt RL, Solomon SC, Grard R, Novara M, and Mukai T (2004) An international program for Mercury exploration: Synergy of MESSENGER and BepiColombo. *Advances in Space Research* 33: 2126–2132.
- McSween HY, Jr (1994) What we have learned about Mars from SNC meteorites. *Meteoritics* 29: 757–779.
- Mellon MT and Jakosky BM (1993) Geographic variations in the thermal and diffusive stability of ground ice on Mars. *Journal of Geophysical Research* 98: 3345–3364.
- Mellon MT and Jakosky BM (1995) The distribution and behavior of Martian ground ice during past and present epochs. *Journal of Geophysical Research* 100: 11781–11800.
- Mellon MT and Phillips RJ (2001) Recent gullies on Mars and the source of liquid water. *Journal of Geophysical Research* 106: 23165–23180.
- Menzies J (ed.) (1995) *Modern Glacial Environments – Processes, Dynamic and Sediments*, 621pp. Oxford, UK: Butterworth-Heinemann.
- Meunier A (2005) *Clays*, 472pp. Berlin, Germany: Springer.
- Mierdel K, Keppler H, Smyth JR, and Langenhorst F (2007) Water solubility in aluminous orthopyroxene and the origin of Earth's asthenosphere. *Science* 315: 364–368.
- Milkovich SM, Head JW, and Marchant DR (2006) Debris-covered piedmont glaciers along the northwest flank of the Olympus Mons scarp: Evidence for low-latitude ice accumulation during the Late Amazonian of Mars. *Icarus* 181: 388–407.
- Milliken RE, Mustard JF, and Goldsby DL (2003) Viscous flow features on the surface of Mars: Observations from high-resolution Mars Orbiter camera (MOC) images. *Journal of Geophysical Research* 108: 5057 (doi:10.1029/2002JE002005).
- Milton DJ (1974) Carbon dioxide hydrate and floods on Mars. *Science* 183: 654–655.
- Mitrofanov I, Antifimov D, Kozyrev A, et al. (2002) Maps of subsurface hydrogen from the high energy neutron detector, Mars Odyssey. *Science* 297: 78–81.
- Mitrofanov IG, Zuber MT, Litvak ML, et al. (2003) CO₂ snow depth and subsurface water-ice abundance in the northern hemisphere of Mars. *Science* 300: 2081–2084.
- Montmessin F (2006) The orbital forcing of climate changes on Mars. *Space Science Review* 125: 457–472 (doi: 10.1007/s11214-006-9078-x).
- Morgenstern A, Hauber E, Reiss D, van Gasselt S, Grosse G, and Schirrmeyer L (2006) Deposition and degradation of a volatile-rich layer in Utopia Planitia, Mars. *Fourth Mars Polar Science Conference*, 8037 (CD-ROM).
- Moses JI, Rawlins K, Zahnle K, and Dones L (1999) External sources of water for Mercury's putative ice deposits. *Icarus* 137: 197–221.
- Mouginis-Mark P (1979) Martian fluidized crater morphology – Variations with crater size, latitude, altitude, and target material. *Journal of Geophysical Research* 84: 8011–8022.
- Mouginis-Mark PJ, Wilson L, Head JW, Brown SH, Hall JL, and Sullivan KD (1984) Elysium Planitia, Mars – Regional geology,

- volcanology, and evidence for volcano–ground ice interactions. *Earth, Moon and Planets* 30: 149–173.
- Murase T (1962) Viscosity and related properties of volcanic rocks at 800° to 1400°C. *Journal of Faculty of Science* pp. 487–584. Hokkaido University, Series 7, geophysics.
- Murchie S, Mustard J, Bishop J, Head J, Pieters C, and Erard S (1993) Spatial variations in the spectral properties of bright regions on Mars. *Icarus* 105: 454–468.
- Murchie SL, Arvidson RE, Bedini P, et al. (2004) CRISM (compact reconnaissance imaging spectrometer for Mars) on MRO (Mars Reconnaissance Orbiter). In: Nardell CA, Lucey PG, Yee JH, and Garvin JB (eds.) *Proceedings of SPIE Vol. 5660: Instruments, Science, and Methods for Geospace and Planetary Remote Sensing*, pp. 66–77. Bellingham, WA: SPIE.
- Murphy DM and Koop T (2005) Review of the vapour pressures of ice and supercooled water for atmospheric applications. *Quaternary Journal of the Royal Meteorological Society* 608: 1539–1565.
- Murray JB, Muller J, Neukum G, et al. (2005) Evidence from the Mars Express high resolution stereo camera for a frozen sea close to Mars' equator. *Nature* 434: 352–356.
- Musselwhite DS, Swindle TD, and Lunine JI (2001) Liquid CO₂ breakout and the formation of recent small gullies on Mars. *Geophysical Research Letters* 28: 1283–1286.
- Mustard JF, Cooper CD, and Riffkin MK (2001) Evidence for recent climate change on Mars from the identification of youthful near-surface ground ice. *Nature* 412: 411–414.
- Neukum G and Hiller K (1981) Martian ages. *Journal of Geophysical Research* 86: 3097–3121.
- Newkirk G, Jr (1981) Solar variability on time scales of 10⁵ years to 10^{9.6} years. *Geochimica et Cosmochimica Acta Supplement* 13: 293–301.
- Newsom H (2005) Clays in the history of Mars. *Nature* 438: 570–571.
- Newsom HE, Brittelle GE, Hibbitts CA, Crossey LJ, and Kudo AM (1996) Impact crater lakes on Mars. *Journal of Geophysical Research* 101: 14951–14956.
- Nozette S, Lichtenberg CL, Spudis P, et al. (1996) The Clementine bistatic radar experiment. *Science* 274: 1495–1498.
- Nozette S, Spudis PD, Robinson MS, Bussey DBJ, Lichtenberg C, and Bonner R (2001) Integration of lunar polar remote-sensing data sets: Evidence for ice at the lunar south pole. *Journal of Geophysical Research* 106: 23253–23266.
- Nummedal D (1978) The role of liquefaction in channel development on Mars. *Reports of Planetary Geology Program*, NASA TM-79729, pp. 257–259.
- Nummedal D and Prior DB (1981) Generation of Martian chaos and channels by debris flows. *Icarus* 45: 77–86.
- Olhoeft GR (1981) Electrical properties of rocks. In: Toulokian YS, Judd WR, and Roy RF (eds.) *Physical Properties of Rocks and Minerals*, pp. 257–330. New York: McGraw-Hill.
- Ori GG, Marinangeli L, and Baliva A (2000a) Terraces in Gilbert-type deltas in crater lakes in Ismenius Lacus and Memnonia (Mars). *Journal of Geophysical Research* 105: 17629–17643.
- Ori G, Marinangeli L, and Komatsu G (2000b) Martian paleolacustrine environments and their geological constraints on drilling operations for exobiological research. *Planetary and Space Science* 48: 1027–1034.
- Ori GG and Mosangini C (1998) Complex depositional systems in hydrates chaos, Mars: An example of sedimentary process interactions in the Martian hydrological cycle. *Journal of Geophysical Research* 103: 22713–22724.
- Page DP and Murray JB (2006) Stratigraphical and morphological evidence for pingo genesis in the cerberus plains. *Icarus* 183: 46–54.
- Paige DA, Herkenhoff KE, and Murray BC (1990) Mariner 9 observations of the south polar CAP of Mars – Evidence for residual CO₂ frost. *Journal of Geophysical Research* 95: 1319–1335.
- Paige DA, Wood SE, and Vasavada AR (1992) The thermal stability of water ice at the poles of Mercury. *Science* 258: 643–646.
- Parker TJ (1989) Channels and valley networks associated with Argyre Planitia, Mars. *Lunar and Planetary Institute Conference*, pp. 826–827.
- Parker TJ, Gorsline DS, Saunders RS, Pieri DC, and Schneeberger DM (1993) Coastal geomorphology of the Martian northern plains. *Journal of Geophysical Research* 98: 11061–11078.
- Paterson WSB (1994) *The Physics of Glaciers*, 3rd edn., 480pp. Oxford: Pergamon.
- Pechmann JC (1980) The origin of polygonal troughs on the northern plains of Mars. *Icarus* 42: 185–210 (doi:10.1016/0019-1035(80)90071-8).
- Perron JT, Dietrich WE, Howard AD, McKean JA, and Pettinga JR (2003) Ice-driven creep on Martian debris slopes. *Geophysical Research Letters* 30: 1747 (doi:10.1029/2003GL017603).
- Phillips RJ, Zuber MT, Solomon SC, et al. (2001) Ancient geodynamics and global-scale hydrology on Mars. *Science* 291: 2587–2591.
- Picardi G, Plaut JJ, Biccari D, et al. (2005) Radar soundings of the subsurface of Mars. *Science* 310: 1925–1928.
- Pieri D (1976) Distribution of small channels on the Martian Surface. *Icarus* 27: 25–50.
- Pollack JB (1991) Kuiper prize lecture: Present and past climates of the terrestrial planets. *Icarus* 91: 173–198.
- Pollack JB, Kasting JF, Richardson SM, and Poliakoff K (1987) The case for a wet, warm climate on early Mars. *Icarus* 71: 203–224.
- Pollack JB, Roush T, Witteborn F, et al. (1990) Thermal emission spectra of Mars (5.4–10.5 microns) – Evidence for sulfates, carbonates, and hydrates. *Journal of Geophysical Research* 95: 14595–14627.
- Pondrelli M, Baliva A, Di Lorenzo S, Marinangeli L, and Rossi AP (2005) Complex evolution of paleolacustrine systems on Mars: An example from the Holden crater. *Journal of Geophysical Research* 110: E04016 (doi:10.1029/2004JE002335).
- Pondrelli M, Rossi AP, Marinangeli L, et al. (2006) Morphofacies analysis of the Eberswalde crater (Mars). *Lunar and Planetary Science XXXVII*: 1555 (CD-ROM).
- Porsild AE (1938) Earth mounds in unglaciated Arctic northwest America. *Geographical Reviews* 28: 46–58.
- Poulet F, Bibring J-P, Mustard JF, et al. (2005) Phyllosilicates on Mars and implications for early Martian climate. *Nature* 438: 623–627.
- Quantin C, Allemand P, Mangold N, Dromart G, and Delacourt C (2005) Fluvial and lacustrine activity on layered deposits in Melas Chasma, Valles Marineris, Mars. *Journal of Geophysical Research* 110: E12S19 (doi:10.1029/2005JE002440).
- Rabette M, McKay C, Pilewskie P, and Young R (2002) The runaway greenhouse effect on Earth. *Astrobiology Science Conference*, NASA Ames Research Center, Moffett Field, CA, 8 April.
- Reedy RC (1978) Planetary gamma-ray spectroscopy. *Lunar and Planetary Science Conference*, pp. 2961–2984.
- Reiss D and Jaumann R (2003) Recent debris flows on Mars: Seasonal observations of the Russell crater dune field. *Geophysical Research Letters* 30(6): 1321 (doi:10.1029/2002GL016704).

- Reiss D, Hauber E, Michael G, Jaumann R, and Neukum G (2005) Small rampart craters in an equatorial region on Mars: Implications for near-surface water or ice. *Geophysical Research Letters* 32: L10202 (doi:10.1029/2005GL022758).
- Reiss D, van Gasselt S, Hauber E, Michael G, Jaumann R, and Neukum G (2006) Ages of rampart craters in equatorial regions on Mars: Implications for the past and present distribution of ground ice. *Meteoritics and Planetary Science* 41: 1437–1452.
- Reiss D, van Gasselt S, Neukum G, and Jaumann R (2004) Absolute dune ages and implications for the time of formation of gullies in Nirgal Vallis, Mars. *Journal of Geophysical Research* 109: E06007 (doi:10.1029/2004JE002251).
- Richardson MI and Mischner MA (2005) Long-term evolution of transient liquid water on Mars. *Journal of Geophysical Research* 110: E03003 (doi:10.1029/2004JE002367).
- Rieder R, Economou T, Wänke H, et al. (1997) The chemical composition of Martian soil and rocks returned by the mobile alpha proton X-ray spectrometer: Preliminary results from the X-ray mode. *Science* 278: 1771–1774.
- Rieder R, Gellert R, Brückner J, et al. (2003) The new Athena alpha particle X-ray spectrometer for the Mars exploration rovers. *Journal of Geophysical Research* 108: 8066 (doi:10.1029/2003JE002150).
- Rivkin AS, Howell EG, Vilas F, and Lebofsky LA (2002) Hydrated minerals on asteroids: The astronomical record. *NASA STI/Recon Technical Report N*, 2, 46, 798.
- Robinson MS and Tanaka KL (1990) Magnitude of a catastrophic flood event at Kasei Vallis, Mars. *Geology* 18: 902–905.
- Rodriguez JAP, Sasaki S, and Miyamoto H (2003) Nature and hydrological relevance of the Shalbatana complex underground cavernous system. *Geophysical Research Letters* 30: 1304 (doi:10.1029/2002GL016547).
- Rodriguez JAP, Tanaka KL, Miyamoto H, and Sasaki S (2006) Nature and characteristics of the flows that carved the Simud and Tiu outflow channels, Mars. *Geophysical Research Letters* 33: L08S04 (doi:10.1029/2005GL024320).
- Rossbacher LA and Judson S (1981) Ground ice on Mars – Inventory, distribution, and resulting landforms. *Icarus* 45: 39–59.
- Roush TL, Blaney DL, and Singer RB (1993) The surface composition of Mars as inferred from spectroscopic observations. In: Pieters CM and Englert PAJ (eds.) *Remote Geochemical Analysis: Elemental and Mineralogical Composition*, pp. 367–393. Cambridge, UK: Cambridge University Press.
- Ruff SW (2004) Spectral evidence for zeolite in the dust on Mars. *Icarus* 168: 131–143.
- Russell PS and Head JW (2003) Elysium-Utopia flows as megafloods: A model of dike intrusion, cryosphere cracking, and water-sediment release. *Journal of Geophysical Research* 108(E6): 5064 (doi:10.1029/2002JE001995).
- Salisbury JW, D'Aria DM, and Jarosewich E (1991) Midinfrared (2.5–13.5 microns) reflectance spectra of powdered stony meteorites. *Icarus* 92: 280–297.
- Schmincke H-U (2000) *Vulkanismus*, 264pp. Darmstadt, Germany: Wissenschaftliche Buchgesellschaft.
- Schorghofer N and Aharonson O (2005) Stability and exchange of subsurface ice on Mars. *Journal of Geophysical Research* 110: E05003 (doi:10.1029/2004JE002350).
- Schultz PH (1992) Atmospheric effects on ejecta emplacement. *Journal of Geophysical Research* 97: 11623–11662.
- Schultz PH and Gault DE (1979) Atmospheric effects on Martian ejecta emplacement. *Journal of Geophysical Research* 84: 7669–7687.
- Scott DH and Dohm JM (1992) Mars highland channels: An age reassessment. *Lunar and Planetary Science* 23: 1251–1252.
- Scott DH, Dohm JM, and Rice JW (1995) *Geological Survey Miscellaneous Investigation Series Map, I-2461-A: Map Showing Channels and Possible Paleolake Basins*. Flagstaff, AZ: US Geological Survey.
- Scott DH and Tanaka KL (1986) *US Geological Survey Miscellaneous Investigation Series Map, I-1802-A: Geologic map of the western equatorial region of Mars*. Flagstaff, AZ: US Geological Survey.
- Sears DWG and Chittenden JD (2005) On laboratory simulation and the temperature dependence of the evaporation rate of brine on Mars. *Geophysical Research Letters* 32: L23203 (doi:10.1029/2005GL024154).
- Seibert NM and Kargel JS (2001) Small-scale Martian polygonal terrain: Implications for liquid surface water. *Geophysical Research Letters* 28: 899–902.
- Shean DE, Head JW, and Marchant DR (2005) Origin and evolution of a cold-based tropical mountain glacier on Mars: The Pavonis Mons fan-shaped deposit. *Journal of Geophysical Research* 110: 5001 (doi:10.1029/2004JE002360).
- Shinbrot T, Duong NH, Kwan L, and Alvarez MM (2004) Dry granular flows can generate surface features resembling those seen in Martian gullies. *Proceedings of the National Academy of Sciences* 101: 8542–8546.
- Shivola A (1996) *Electromagnetic Mixing Formulas and Applications*, 284pp. London: IEE.
- Simpson RA and Tyler GL (1999) Reanalysis of Clementine bistatic radar data from the lunar South Pole. *Journal of Geophysical Research* 104: 3845–3862.
- Slade MA, Butler BJ, and Muhleman DO (1992) Mercury radar imaging – Evidence for polar ice. *Science* 258: 635–640.
- Smith DE, Zuber MT, Frey HV, et al. (2001) Mars Orbiter laser altimeter: Experiment summary after the first year of global mapping of Mars. *Journal of Geophysical Research* 106: 23689–23722.
- Smith DE, Zuber MT, Solomon SC, et al. (1999) The global topography of Mars and implications for surface evolution. *Science* 284: 1495–1502.
- Soare RJ, Bun Tseung JMW, and Peloquin C (2005) Possible thermokarst and alas formation in Utopia Planitia, Mars. *Lunar and Planetary Science XXXVI*: 1103 (CD-ROM).
- Soderblom LA (1992) The composition and mineralogy of the Martian surface from spectroscopic observations – 0.3 micron to 50 microns. In: Kieffer HH, Jakosky BM, Synder CW, and Matthews MS (eds.) *Mars*, pp. 557–593. Tucson, AZ: University of Arizona Press.
- Soderblom LA, Kreidler TJ, and Masursky H (1973) Latitudinal distribution of a debris mantle on the Martian surface. *Journal of Geophysical Research* 78: 4117–4122.
- Solomon SC, Aharonson O, Aurnou JM, et al. (2005) New perspectives on ancient Mars. *Science* 307: 1214–1220.
- Solomon SC, McNutt RL, Gold RE, et al. (2001) The MESSENGER mission to Mercury: Scientific objectives and implementation. *Planetary and Space Science* 49: 1445–1465.
- Soukhovitskaya V and Manga M (2006) Martian landslides in Valles Marineris: Wet or dry? *Icarus* 180: 348–352 (doi:10.1016/j.icarus.2005.09.008).
- Sparks RSJ, Gardeweg MC, Calder ES, and Matthews SJ (1997) Erosion by pyroclastic flows on Lascar Volcano, Chile. *Bulletin of Volcanology* 58: 557–565.
- Squyres SW (1978) Martian fretted terrain – Flow of erosional debris. *Icarus* 34: 600–613.
- Squyres SW (1979) The distribution of lobate debris aprons and similar flows on Mars. *Journal of Geophysical Research* 84: 8087–8096.
- Squyres SW (1989) Urey prize lecture – Water on Mars. *Icarus* 79: 229–288 (doi:10.1016/0019-1035(89)90078-X).

- Squyres SW, Arvidson RE, Baumgartner ET, et al. (2003) Athena Mars rover science investigation. *Journal of Geophysical Research* 108: 8062 (doi:10.1029/2003JE002121).
- Squyres SW and Carr MH (1986) Geomorphic evidence for the distribution of ground ice on Mars. *Science* 231: 249–252.
- Squyres SW, Clifford SM, Kuzmin RO, Zimbelman JR, and Costard FM (1992) Ice in the Martian regolith. In: Kieffer HH, Jakosky BM, Snyder CM, and Matthews MS (eds.) *Mars*, pp. 523–554. Tucson, AZ: University of Arizona Press.
- Squyres SW, Grotzinger JP, Arvidson RE, et al. (2004) *In situ* evidence for an ancient aqueous environment at Meridiani Planum, Mars. *Science* 306: 1709–1714.
- Squyres SW and Kasting JF (1994) Early Mars – How warm and how wet. *Science* 265: 744–749.
- Squyres SW, Knoll AH, Arvidson RE, et al. (2006) Two years at Meridiani Planum: Results from the Opportunity rover. *Science* 313: 1403–1407.
- Stacey NJS, Campbell DB, and Ford PG (1997) Arecibo radar mapping of the lunar poles: A search for ice deposits. *Science* 276: 1527–1530.
- Stepinski TF and Stepinski AP (2005) Morphology of drainage basins as an indicator of climate on early Mars. *Journal of Geophysical Research* 110: E12S12 (doi:10.1029/2005JE002448).
- Stewart ST and Nimmo F (2002) Surface runoff features on Mars: Testing the carbon dioxide formation hypothesis. *Journal of Geophysical Research* 107: 5069 (doi:10.1029/2000JE001465).
- Stewart ST, O'Keefe JD, and Ahrens TJ (2001) The relationship between Rampart crater morphologies and the amount of subsurface ice. *Lunar and Planetary Science XXXII*: 2092 (CD-ROM).
- Stockstill KR, Moersch JE, Ruff SW, Baldridge A, and Farmer J (2005) Thermal emission spectrometer hyperspectral analyses of proposed paleolake basins on Mars: No evidence for in-place carbonates. *Journal of Geophysical Research* 110: 10004 (doi:10.1029/2004JE002353).
- Sugden DE and John BS (1976) *Glaciers and Landscape*, 328pp. London: Arnold.
- Summerfield MA (1991) *Global Geomorphology*, 535pp. London: Longman.
- Tanaka KL (1997) Sedimentary history and mass flow structures of Chryse and Acidalia Planitiae, Mars. *Journal of Geophysical Research* 102: 4131–4150.
- Tanaka KL (1999) Debris-flow origin for the Simud/Tiu deposit on Mars. *Journal of Geophysical Research* 104: 8637–8652.
- Tanaka KL, Banerdt WB, Kargel JS, and Hoffman N (2001) Huge, CO₂-charged debris-flow deposit and tectonic sagging in the northern plains of Mars. *Geology* 29: 427–430.
- Terada N, Machida S, and Shinagawa H (2002) Global hybrid simulation of the Kelvin–Helmholtz instability at the Venus ionopause. *Journal of Geophysical Research* 107(A12): 1471–1490.
- Theilig E and Greeley R (1979) Plains and channels in the Lunae Planum-Chryse Planitia region of Mars. *Journal of Geophysical Research* 84: 7994–8010.
- Thomas P, Squyres S, Herkenhoff K, Howard A, and Murray B (1992) *Polar Deposits of Mars, in Mars*, pp. 767–795. Tucson, AZ: University of Arizona Press.
- Tokano T and Bish DL (2005) Hydration state and abundance of zeolites on Mars and the water cycle. *Journal of Geophysical Research* 110: E12S08 (doi:10.1029/2005JE002410).
- Tomasko MG, Archinal B, Becker T, et al. (2005) Rain, winds and haze during the Huygens probe's descent to Titan's surface. *Nature* 438: 765–778.
- Toon OB, Pollack JB, Ward W, Burns JA, and Bilski K (1980) The astronomical theory of climatic change on Mars. *Icarus* 44: 552–607.
- Toulmin P, Rose HJ, Christian RP, et al. (1977) Geochemical and mineralogical interpretation of the Viking inorganic chemical results. *Journal of Geophysical Research* 82: 4625–4634.
- Touma J and Wisdom J (1993) The chaotic obliquity of Mars. *Science* 259: 1294–1297.
- Treiman AH (2003) Geologic settings of Martian gullies: Implications for their origins. *Journal of Geophysical Research* 108: 8031 (doi:10.1029/2002JE001900).
- Twidale CR (1976) *Analysis of Landforms*, 572pp. Sydney, NSW: John Wiley.
- Urey HC (1967) Water on the Moon. *Nature* 216: 1094.
- Vallance JW (2000) Lahars. In: Sigurdsson H (ed.) *Encyclopedia of Volcanoes*, pp. 601–616. San Diego, CA: Academic Press.
- van Everdingen RO (2005) Mult-Language Glossary of Permafrost and Related Ground-Ice Terms, International Parmafrost Association, Caalgary, <http://nsidc.org/fgdc/glossary>, (accessed on Feb 2007).
- Van Gasselt S, Hauber E, and Neukum G (in press) Cold-climate modification of Martian landscapes: A case study of a spatuolate debris landform in the Hellas Montes Region, Mars. *Journal of Geophysical Research*.
- van Gasselt S, Reiss D, Thorpe AK, and Neukum G (2005) Seasonal variations of polygonal thermal contraction crack patterns in a south polar trough, Mars. *Journal of Geophysical Research (Planets)* 110: 8002 (doi:10.1029/2004JE002385).
- van Steijn P (1989) Puinstromen. *Geografisch Tijdschrift* 23: 119–128.
- Vaniman DT, Chipera SJ, Bish DL, Carey JW, Fialips CL, and Feldman WC (2004) Sulfate salts, regolith interactions, and water storage in equatorial Martian regolith. *Lunar and Planetary Science XXXV*: 1690 (CD-ROM).
- Vasavada AR, Paige DA, and Wood SE (1999) Near-surface temperatures on Mercury and the moon and the stability of polar ice deposits. *Icarus* 141: 179–193.
- Vialov SS (1958) Regularities of glacial shields movements and the theory of plastic viscous flow. *Physics of the Movements of Ice, International Association of Hydraulic Sciences* 47: 266–275.
- Wahrhaftig C and Cox A (1959) Rock glaciers in the Alaska Range. *Geological Society of America, Bulletin* 70: 383–436.
- Wallace P and Anderson A (2000) Volatiles in magmas. In: Sigurdsson H (ed.) *Encyclopedia of Volcanoes*, pp. 149–170. New York: Academic Press.
- Wang A, Korotev RL, Jolliff BL, et al. (2006) Evidence of phyllosilicates in Wooly Patch, an altered rock encountered at West Spur, Columbia Hills, by the Spirit rover in Gusev crater, Mars. *Journal of Geophysical Research* 111: E02S16 (doi:10.1029/2005JE002516).
- Wänke H, Brückner J, Dreibus G, Rieder R, and Ryabchikov I (2001) Chemical composition of rocks and soils at the pathfinder site. *Space Science Reviews* 96: 317–330.
- Ward WR (1973) Large-scale variations in the obliquity of Mars. *Science* 181: 260–262.
- Ward WR (1974) Climatic variations on Mars. I: Astronomical theory of insolation. *Journal of Geophysical Research* 79: 3375–3386.
- Washburn AL (1979) *Geocryology*, 2nd edn., 406pp. London: Arnold.
- Watson K, Murray BC, and Brown H (1961) The behavior of volatiles on the lunar surface. *Journal of Geophysical Research* 66: 3033–3045.

- Webb VE (2004) Putative shorelines in northern Arabia Terra, Mars. *Journal of Geophysical Research* 109: 9010 (doi:10.1029/2003JE002205).
- Weitz CM, Irwin RP, Chuang FC, Bourke MC, and Crown DA (2006) Formation of a terraced fan deposit in Coprates Catena, Mars. *Icarus* 184: 436–451.
- Werner SC (2006) Major aspects of the chronostratigraphy and geologic evolutionary history of Mars. Dissertation, Freie Universität Berlin, 160 S.
- Werner SC, van Gasselt S, and Neukum G (2003) Continual geological activity in Athabasca Valles, Mars. *Journal of Geophysical Research* 108: 8081 (doi:10.1029/2002JE002020).
- Whalley WB and Azizi F (2003) Rock glaciers and protalus landforms: Analogous forms and ice sources on Earth and Mars. *Journal of Geophysical Research* 108: 8032 (doi:10.1029/2002JE001864).
- Wharton RA, Crosby JM, McKay CP, and Rice JW (1995) Paleolakes on Mars. *Journal of Paleolimnology* 13: 267–283.
- Williams H and Mc Birney AR (1979) *Volcanology*, 397pp. San Francisco, CA: Freeman & Cooper.
- Williams RM, Phillips RJ, and Malin MC (2000) Flow rates and duration within Kasei Valles, Mars: Implications for the formation of a Martian ocean. *Geophysical Research Letters* 27: 1073–1076.
- Williams RME and Phillips RJ (2001) Morphometric measurements of Martian valley networks from Mars Orbiter laser altimeter (MOLA) data. *Journal of Geophysical Research* 106: 23737–23752.
- Williams RS (1978) Geomorphoc processes in Iceland and on Mars: A comparative appraisal from orbital images. *Geological Society of America Abstract Programs* 10: 517.
- Wilson L, Ghatala GJ, Head JW, and Mitchell KL (2004) Mars outflow channels: A reappraisal of the estimation of water flow velocities from water depths, regional slopes, and channel floor properties. *Journal of Geophysical Research* 109: 9003 (doi:10.1029/2004JE002281).
- Woods AW (1995) The dynamics of explosive volcanic eruptions. *Reviews of Geophysics* 33: 495–530.
- Wood BE, Muller HR, Zank G, and Linsky JL (2002) Measured mass loss rates of solar like stars as a function of age and activity. *Astrophysical Journal* 574: 412–425.
- Woronow A (1981) Preflow stresses in Martian rampart ejecta blankets – A means of estimating the water content. *Icarus* 45: 320–330.
- Wyatt MB and McSween HY (2002) Spectral evidence for weathered basalt as an alternative to andesite in the northern lowlands of Mars. *Nature* 417: 263–266.
- Yershov ED (1998) *General Geocryology*, 580pp. Cambridge, UK: Cambridge University Press.
- Yoshikawa K (2003) Origin of the polygons and the thickness of Vastitas Borealis Formation in Western Utopia Planitia on Mars. *Geophysical Research Letters* 30(12): 1603 (doi:10.1029/2003GL017165).
- Yung YL and Pinto JP (1978) Primitive atmosphere and implications for the formation of channels on Mars. *Nature* 288: 735–738.
- Zahnle KJ and Walker JCG (1982) The evolution of solar ultraviolet luminosity. *Reviews of Geophysics and Space Physics* 20: 280–292.
- Zent AP (1999) An open, snow-based, hydrologic system as an analog for Noachian Mars. *Lunar and Planetary Science XXX*: 1803 (CD-ROM).
- Zent AP, Howard DJ, and Quinn RC (2001) H₂O adsorption on smectites: Application to the diurnal variation of H₂O in the Martian atmosphere. *Journal of Geophysical Research* 106: 14667–14674.
- Zent AP and Quinn RC (1995) Simultaneous adsorption of CO₂ and H₂O under Marslike conditions and application to the evolution of the Martian climate. *Journal of Geophysical Research* 100: 5341–5349.
- Zent AP and Quinn RC (1997) Measurement of H₂O adsorption under Mars-like conditions: Effects of adsorbent heterogeneity. *Journal of Geophysical Research* 102: 9085–9096.
- Zhang MHG, Luhmann JG, Bouger SW, and Nagy AF (1993) The ancient oxygen exosphere of Mars implication for atmospheric evaluations. *Journal of Geophysical Research* 98: 10915–10923.
- Zimbelman JR (2001) Image resolution and evaluation of genetic hypotheses for planetary landscapes. *Geomorphology* 37: 179–199.

10.12 Geology, Life and Habitability

G. Southam, The University of Western Ontario, London, ON, Canada

F. Westall, CNRS, Orleans, France

© 2007 Elsevier B.V. All rights reserved.

10.12.1	Introduction	421
10.12.1.1	The Requirements for an Origin of Life and ‘Habitability’	421
10.12.1.2	The Origin of Life and the First Cells	422
10.12.2	Geology, Life and Habitability	423
10.12.2.1	The Habitability Limits: Physical and Chemical Constraints on Life	423
10.12.2.2	Building Cells	425
10.12.2.3	Generating Energy	426
10.12.2.4	The Growth of Microorganisms	427
10.12.3	Geology and Life	429
10.12.3.1	Lithotrophy	429
10.12.3.2	Microorganisms and Mineral Formation	429
10.12.3.3	Habitable Conditions on Early Earth (and Other Early Rocky Planets)	431
10.12.3.4	The Earliest Traces of Life	431
10.12.4	Conclusions for Geology, Life and Habitability beyond the Earth	432
References		433

10.12.1 Introduction

10.12.1.1 The Requirements for an Origin of Life and ‘Habitability’

Whether or not life originated on Earth, it was certainly present in abundance and at a relatively advanced level of evolution by about 3.5 billion years ago (Ga) (see review in Westall and Southam (2006)). At this time, the habitable environment of the Earth was very much different than the Earth of today – significantly hotter (Arndt, 1994; Knauth and Lowe, 2003) and more volcanically/hydrothermally active, and basically without oxygen (<0.2% present atmospheric levels; Kasting, 1993; Catling and Claire, 2005). The low levels of reactive oxygen produced by photolysis (Kasting, 1993) relative to the abundance of reduced chemical species would have resulted in a correspondingly reducing chemistry for the hydrosphere and lithosphere as well. The environment was also affected by a significantly higher bolide flux in the first 750 My of Earth’s history (up to about 3.85 Ga) that could potentially have evaporated all the surface water (Sleep *et al.*, 1989). It is commonly believed that the life on Earth today evolved from primitive cells that could only have appeared after the end of the late heavy bombardment (Lowe and Byerly, 1986; Lowe *et al.*, 2003), that is, after ~3.85 Ga. However, the last

decade has seen an enormous advance in our understanding of the limits of life and it is clear that, once started, it is very difficult to eradicate prokaryotes from either the Bacteria or Archaea domains (Woese *et al.*, 1990), especially the latter since they generally occur in environments that today are considered to be extreme but were common on the early Earth. It is possible, therefore, that life appeared on Earth before ~4.0 Ga and survived the catastrophic bolide flux in, for instance, a protected subsurface habitat, to reappear and flourish at the surface once water had re-condensed. This hypothesis is supported to a certain extent by the fact that gene sequence analysis points to hyperthermophilic microorganisms as being at the base of the evolutionary tree (Pace, 1997). It is not believed that the hyperthermophiles are the oldest living forms of life, rather they are thought to represent a ‘hot bottleneck’ through which the early forms of life had to pass, such as would have existed during a catastrophic impact event. An alternative hypothesis is that life first appeared on another planet and was transported by panspermia to the Earth once habitable conditions had been re-established at the end of the catastrophic bolide period. However, the other likely contenders for an independent origin of life, Mars, and maybe Venus, also suffered from the same bolide bombardment (see Chapter 10.06).

In order to address the concept of life and habitability in the solar system, we start with a brief overview of current thinking on the basic requirements for life to originate and the origin of life. In life as we know it, all cells need a source of energy, and are composed of water and carbon molecules that include some essential elements (notably hydrogen, oxygen, nitrogen, and sulfur; HONS). These are therefore the minimum requirements on a planet (or other type of rocky body) for the appearance of life. Since organic carbon molecules occur in interstellar clouds, as well as in the materials forming the proto solar system cloud (in the form of the planetesimals that formed the rocky planets, asteroids, meteorites, micrometeorites, and comets), they can be considered to be more or less ubiquitous. Endogenous sources of organic carbon (Miller, 1953) are primarily produced in Fischer Tropsch synthesis in hydrothermal systems (e.g., hydrocarbons containing up to 29 carbon atoms have been produced from the Rainbow ultramafic hydrothermal system on the mid-Atlantic ridge; Holm and Charlot, 2001), which are believed to have been very active on the early Earth. A significant amount of the organic molecules (and other volatiles, such as water) were also delivered from extraterrestrial sources (Brown *et al.*, 2000). Carbonaceous chondrite meteorites contain up to 5% carbon and eight of the amino acids that have been identified in them are of the kind used in living cells today; they are the building blocks of life. The present day flux of carbonaceous micrometeorites containing up to 2% carbon to the Earth is estimated at 50–100 tons yr⁻¹, which implies an extraordinary amount of 10²³ g carbon over a 300 My period during the late heavy bombardment (Maurette *et al.*, 2000; see also Brack (2004) for a review). Thus, organic carbon on rocky bodies in the young solar system was readily available.

The sources of the other elements essential for an origin of life (HONS) would likewise be readily available on any rocky planet. A rocky planet is also an excellent source of energy. Living organisms need energy in order to mitigate against the constant problem of entropy their entire lives. From a simplistic, energy perspective then, any geologic system that provides geochemical energy, that is, a reaction that can be written possessing a negative ΔG_r should conceivably have the capacity to support life. However, the occurrence of life on Earth and the potential for life to occur elsewhere in our solar system (or in the universe) requires that the target planetary body falls within the habitability zone for the planetary system (see Franck *et al.* (2006)

and Kasting *et al.* (1993)). The habitability zone is defined by the sum of the physical and chemical conditions, which would support the presence of liquid water. The habitability zone around the Sun has moved outward since the formation of the solar system as luminosity, which is now about 25–30% greater than 4.6 Ga has increased (Newman and Rood, 1977). Rocky planets within this zone represent ideal potential habitats. Rocky bodies with liquid water in contact with the rock do exist outside this zone and can thus be considered as being potentially habitable. In the early part of the history of the solar system, Europa had its own internal heat source from decaying short-lived radioactive elements. It is believed that there is still a liquid ocean underneath its icy crust, the energy required to keep the fluid liquid being imparted by the very strong tidal forces from Jupiter (Chyba and Phillips, 2001).

10.12.1.2 The Origin of Life and the First Cells

A definition of a minimal life form could be “a structure capable of auto-reproduction including errors that lead to evolution” (Brack, 2005). The basic structure consists of three long chain carbon molecules that are necessary for the functioning of the cell: a lipid membrane containing a molecule capable of information transfer (RNA/DNA) and a catalyzing molecule (enzyme) to provide the basic work of the cell. Despite the fact that it has not yet been possible to reproduce a life form *ex novo* in the laboratory, the last 50 years has seen significant advances in understanding how these basic elements formed from inorganic components.

In the modern cell, membranes are primarily composed of amphipathic phospholipids that are stable in water. These lipids can be formed from fatty acids, which occur in carbonaceous chondrites like the Murchison meteorite (Deamer, 1998). Catalytic reactions in modern, living cells are performed by proteinaceous enzymes, themselves made of homochiral amino acids. As we have noted above, these were abundantly available on the early Earth. There are, however, a number of scenarios regarding the condensation or assemblage of simple amino acids into longer chain molecules, such as peptides. These include condensation on mineral surfaces, such as clays (Ferris *et al.*, 1996); thermal condensation (Fox and Dose, 1977); and chemical reaction (Barbier and Brack, 2002). Similarly, the origin of the first information transfer molecules is highly

debated. In modern cells the nucleic acids are composed of bases (purine and pyrimidine), sugars (ribose or deoxyribose), and phosphate groups. RNA is considered to be a possible, primitive replicating molecule because it also has catalytic properties (representing self-replication, mutation, and evolution) and can act as polymerization templates (Zaug and Cech, 1986). However, even RNA is too complicated to have been the first replicating molecule (Schwartz and Orgel, 1985). Nonenzymatic replication has been demonstrated (Inoue and Orgel, 1982), as have autocatalytic reactions involving the reaction of FeS and H₂S on the surface of pyrite (Wächtershäuser, 1988). In the latter scenario, there is an evolution of organic molecules forming a network on the surface of the mineral. Other candidates for precursor replication molecules include pyranosyl-RNA (p-RNA) and threo furanosyl nucleic acid (TNA), which is more likely because of the greater availability of the carbon precursors of tetrose than ribose on the primitive Earth.

The basic process leading to an origin of life can be thus envisaged (Brack, 2004): CHONS + water → catalysts → RNA world → cells (RNA, proteins and membranes).

10.12.2 Geology, Life and Habitability

Based on the above-described criteria for habitability and the origin of life, the key to understanding the current, potential or historical interaction between geology and life, that is, habitability, is the understanding of the history of liquid water on any terrestrial planet or icy moon. Given the necessity of disequilibria conditions, the liquid water needs to be in contact with minerals/rocks, implying liquid water on a rocky planet. The capacity to support life may be transient, that is, it will not be constant for a particular planet over time but will vary based on a wide range of geologic and planetary factors. From the life perspective, we only have one life form to use as a basis to discuss the relationship between geology (including aqueous geochemistry), life and habitability. Assuming water is available, there is a wide range or diversity of metabolic strategies that have been developed by terrestrial microorganisms (Barns and Nierwizcki-Bauer, 1997; Pace, 1997; Reysenbach and Shock, 2002), so that they can successfully colonize geologic environments.

In this chapter, we focus on the interaction between geology and prokaryotic life as a way to address the subject of habitability, both from our perspective and from that of the prokaryotes, which have inhabited Earth for far longer than we have. Today on Earth, prokaryotic organisms are ubiquitous, being found in a wide variety of surface and subsurface environments (Colwell *et al.*, 1997; Lovley and Chapelle, 1995; Parkes *et al.*, 1994; Stevens *et al.*, 1993), where they are only limited by the availability of water (Navarro-Gonzalez *et al.*, 2003) and temperatures less than 121°C (Kashefi and Lovley, 2003). In these natural environments, prokaryotes typically grow as biofilms on mineral surfaces (Marshall, 1988; Wanger *et al.*, 2006). This intimate association of life with its geological substrate is thought to have occurred for the last ~3.5 billion years (Furnes *et al.*, 2004; see Westall and Southam (2006), for a review) producing molecular to global scale byproducts (Brocks *et al.*, 1999; Konhauser *et al.*, 2002; Newman and Banfield, 2002). During the initial stages of this interaction, the physical and chemical conditions of the early Earth would have controlled the biosphere. However, over time, life exerted an increasing influence over the geosphere. Life serves as a catalyst for a wide array of reactions at geologically low temperatures, doing them faster than would happen abiotically (e.g., Singer and Stumm, 1970).

10.12.2.1 The Habitability Limits: Physical and Chemical Constraints on Life

The growth of prokaryotes is affected by the various physical and chemical constraints in the environments in which they are growing. They, in turn, can have a profound effect on their surrounding environment. Understanding the physical and chemical limits of bacteria and their environmental influences allows us to predict the types and distribution of microorganisms in natural systems.

The range of (and optimum) growth temperatures differentiates prokaryotes into broad groups (Stetter *et al.*, 1990). Psychrophiles grow optimally at or below 15°C, and have even demonstrated activity down to minus 20°C; viable prokaryotes have been recovered from million-year-old permafrost (Rivkina *et al.*, 2000). Psychrotolerant prokaryotes grow optimally between 20°C and 40°C but will grow at temperatures as low as 0°C. Mesophiles grow optimally between 15°C and 45°C. Thermophiles have optimum growth temperatures above 45°C and hyperthermophiles grow optimally above 80°C



Figure 1 A photograph of a hydrothermal spring in Norris Geyser Basin, Yellowstone National Park. Extreme environments, such as this, are dominated by prokaryotic organisms representing both the Archaea and Bacteria domains. The hydrothermal feature is ~8 m across.

(**Figure 1**). While no single microorganism can withstand the full range of temperatures described here, as a group, prokaryotes can be found across any thermal gradient, from ice or permafrost regions all the way up to hydrothermal fluids (121°C).

In all environments, prokaryotes concentrate solutes within their cell envelopes and obtain water through osmosis, thus enabling them to maintain intracellular turgor pressure (Kunte, 2006). The prokaryote cell envelope is designed to withstand this turgor pressure which can measure between 2–3 atm in Gram-negative bacteria and up to 15 atm in Gram-positive bacteria (Beveridge, 1981). The accumulation of water represents a challenge for prokaryotes (Archaea) in contemporary- (e.g., the Great Salt Lake) and paleo- (Bechtel *et al.*, 1996) evaporitic, hypersaline systems. Some of these environments are continually under stress due to low water activity and are therefore considered to be examples of extreme environments, which will only support halophilic (or halotolerant) microorganisms (see **Figure 2**).

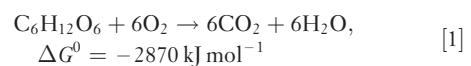
Most natural environments occur as circumneutral pH systems. At pH 7 the hydrogen ion concentration is 100 nM. Therefore, most prokaryotes exhibit optimal activity in the pH range of 6–8. Extreme environments possessing acidic (e.g., pH 2) or alkaline (e.g., pH 11) pH conditions will contain indigenous microorganisms, that is, acidophilic (acidotolerant) or alkaliphilic (alkalitolerant) prokaryotes, respectively, whose growth parameters reflect, that is, are adapted to, these extremes in



Figure 2 A photograph of Basque Lake #2, in central BC, Canada. The evaporative minerals in this basin are dominated by epsomite, indicating the hypersaline nature of this environment. The hypersaline pool is ~5 m in diameter.

environmental pH (Goodwin and Zeikus, 1987). Note, the word ‘extremophiles’ is an anthropocentric term, for example, for many thiobacilli, pH 2 is optimum or ideal (ambient).

Our modern habitat is characterized by an atmosphere rich in oxygen (which was not the case on the early Earth). The capacity to utilize oxygen as a terminal electron acceptor creates a competitive advantage and a challenge for prokaryotes, which are subdivided into aerobic and anaerobic groups based on their oxygen requirements and detoxification strategies (Chapelle, 1993). There are also facultatively anaerobic prokaryotes that utilize oxygen if it is available, becoming anaerobes when the oxygen runs out, that is, growth with or without oxygen. The utilization of oxygen as a terminal electron acceptor enables organisms to fully oxidize their respective source(s) of energy and thereby maximize their energy gain from the reaction (reaction [1]):



Any organism which is able to grow in the presence of oxygen, whether or not it utilizes oxygen as a terminal electron acceptor, must also have the capacity to detoxify the chemically reactive oxygen compounds produced through the stepwise reduction of oxygen to water (hydrogen peroxide, superoxide radical, and the hydroxyl radical; Atlas and Bartha, 1997). Detoxification of these oxygen radicals requires a series of enzymes including superoxide dismutase, catalase, and peroxidase. The ability to tolerate oxygen relates to the presence or absence



Figure 3 A photograph of a hydrothermal acid hot spring in Shimabara, Japan. In these extreme environments, the geosphere dominates the biosphere.

of each of these enzymes. Measurements of dissolved oxygen and redox potential, which represents the sum reactivity of oxidants and reductants, are important in determining which microbial processes are functioning within natural systems.

Analyses of the physical and chemical limits on prokaryote growth and survival typically focus on one ‘stressor’ at a time, for example, pH (acidophile vs neutrophile vs alkaliphile), temperature

(psychrophile vs mesophile vs thermophile), salinity (freshwater vs halophile) or oxygen availability (aerobe vs anaerobe). However, there are many examples of prokaryotes that can withstand multiple ‘stressors’, for example, thermoacidophiles. It is in these environments (see **Figure 3**) that the diversity of prokaryotes and the limits of life are truly appreciated (see **Table 1**). However, other combinations of ‘stressors’ will be deadly to microorganisms, such as low water activity together with high UV radiation, as on the surface of Mars today (Cockell and Raven, 2004).

10.12.2.2 Building Cells

Modern prokaryotes use a wide range of organic and inorganic nutrients. Organic nutrients, for example, carbohydrates, amino acids, nucleic acids, and hydrocarbons, include nearly every organic monomer or polymer found in the living or diagenetically altered biosphere. Inorganic nutrients include dissolved gases (carbon monoxide, carbon dioxide, hydrogen, nitrogen, and dihydrogen sulfide), soluble cations (sodium, calcium, magnesium, ammonium, ferrous, and ferric iron), base metals (chromium, nickel, copper, cobalt, zinc, lead, mercury), and soluble anions (chloride, nitrite, nitrate, hydrogen sulfide, sulfite,

Table 1 Classification and examples of extremophiles

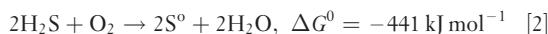
Environmental parameter	Type	Definition	Examples
Temperature	Hyperthermophile	Growth > 80°C	<i>Pyrolobus fumarii</i> , 113°C
	Thermophile	Growth 60–80°C	<i>Synechococcus lividis</i>
	Mesophile	15–60°C	<i>Homo sapiens</i>
	Psychrophile	<15°C	<i>Psychrobacter</i> , some insects
Radiation	Barophile	Weight loving	<i>Deinococcus radiodurans</i>
	Piezophile	Pressure loving	Unknown
Gravity	Hypergravity	>1 g	For microbe, 130 MPa
	Hypogravity	<1 g	None known
Vacuum		Tolerates vacuum (space devoid of matter)	Tardigrades, insects, microbes, seeds
Desiccation	Xerophiles	Anhydrobiotic	<i>Artemia salina</i> ; nematodes, microbes, fungi, lichens
Salinity	Halophile	Salt loving (2–5 M NaCl)	<i>Halobacteriaceae</i> , <i>Dunaliella salina</i>
pH	Alkaliphile	pH > 9	<i>Natronobacterium</i> , <i>Bacillus firmus</i> OF4, <i>Spirulina</i> spp. (all pH 10.5)
Oxygen tension	Acidophile	Low pH loving	<i>Cyanidium caldarium</i> , <i>Ferroplasma</i> sp. (both pH 0)
	Anaerobe	Cannot tolerate O ₂	<i>Methanococcus jannaschii</i>
	microaerophile	Tolerates some O ₂	<i>Clostridium</i>
Chemical extremes	aerobe	requires O ₂	<i>Homo sapiens</i>
	Gases metals	Can tolerate high concentrations of metal (metalotolerant)	<i>Cyanidium caldarium</i> (pure CO ₂) <i>Ferroplasma acidarmanus</i> (Cu, As, Cd, Zn); <i>Ralstonia</i> sp. CH34 (Zn, Co, Cd, Hg, Pb)

From Rothschild and Mancinelli (2001).

sulfate, phosphorus, selenate, and arsenate). Depending on nutrient requirements, these compounds can be used in assimilation reactions, that is, building new cells, or in dissimilatory processes to generate the energy needed to construct new biomass (discussed below).

Carbon utilization is one of the most important criteria by which prokaryotes are characterized. Heterotrophic prokaryotes require organic carbon for their metabolism and typically couple the oxidation of organic carbon to CO_2 with the reduction of dissolved inorganic species or minerals. The sources of carbon and energy for heterotrophs can range from simple organics such as glucose or asparagine, an amino acid (Goldman and Wilson, 1977), to macromolecular materials such as cellulose (Ljungdahl and Eriksson, 1985).

Autotrophs are able to obtain their cellular carbon for biomass from inorganic sources (dissolved $\text{CO}_2/\text{HCO}_3^-$). Autotrophic organisms direct most of their energy toward the fixation (reduction) of CO_2 into organic carbon, that is, biomass production, and therefore tend to grow at slower rates than heterotrophic organisms. However, autotrophs are extremely important since they serve as the basis of the food chain upon which ‘we’ heterotrophs thrive. For example, autotrophic sulfur oxidizing bacteria (reaction [2]) that synthesize carbohydrates at seafloor hot springs (reaction [3]):



serve as the base of the ‘food chain’ in these lightless, extreme environments (Cavanaugh *et al.*, 1981).

Nitrogen and phosphorus are limiting nutrients in most natural systems. However, prokaryotes can easily assimilate all forms of water-soluble inorganic nitrogen and when this nitrogen is limiting, a wide range of prokaryotes possess nitrogenase activity and can fix atmospheric nitrogen (Brill, 1975). Regarding phosphorus limiting conditions, microorganisms can produce a wide array of organic and inorganic acids that can enhance silicate weathering and phosphate assimilation (Rogers *et al.* (1998); discussed below).

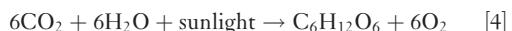
10.12.2.3 Generating Energy

In natural systems, prokaryotes exploit a wide range of redox reactions that possess negative ΔG_r to support metabolism (Nealson and Stahl, 1997). The requirement of energy for biosynthetic reactions is

described according to the electron donor or source of reducing power and by the electron acceptor, that is, a redox couple. Current research on the diversity of life is focusing on phenotypic and genotypic studies of geochemical gradients, where thermodynamic calculations highlight possible biogeochemical reactions that could support life, for example, anaerobic methane oxidation at deep sea clathrate vents (Thomsen *et al.*, 2001).

In natural systems, bacterially catalyzed redox processes are established through competition for available nutrients (assimilation) and through the efficiency of their respective energy generation mechanisms (Lovley and Klug, 1983). As geochemical conditions become increasingly reducing, the amount of energy (as expressed by ΔG_r) available from each of the predominant biogeochemical reactions (i.e., redox zones) decreases (Lovley and Phillips, 1988; Lovley and Chapelle, 1995) resulting in a concomitant reduction in biomass (Onstott *et al.*, 2006).

Phototrophs utilize sunlight to produce adenosine triphosphate (ATP, the principal energy carrier of the cell) from otherwise chemically stable reduced compounds. Anoxygenic phototrophs were probably already present on the Earth by 3.5 Ga (see review in Westall and Southam (2006)); contemporary representatives of this group are anaerobic green and purple sulfur bacteria that derive their energy for growth from light and H_2S (or Fe^{2+} ; Heising and Schink (1998) and Kappler and Newman (2004)), producing elemental sulfur or sulfate (or Fe^{3+}) as end products of their metabolism (van Gemerden, 1986). The later-evolved oxygenic phototrophs use light in combination with H_2O as their source of energy (reaction [4]):



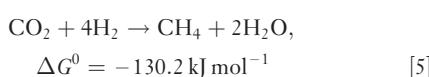
Chemolithotrophs (hydrogen oxidizers, ammonia oxidizers, sulfur oxidizers, and methylotrophs) are prokaryotes which utilize inorganic chemicals (H_2 , NH_4^+ , H_2S , and CH_4) as their energy sources (electron donors). Lithotrophs grow more slowly than do heterotrophs owing to the relatively low levels of chemical energy present in inorganic versus organic compounds (Nealson and Stahl, 1997). Chemoorganotrophs, which utilize organic carbon as their source of reducing power and cellular carbon, include both aerobic and anaerobic prokaryotes.

Prior to the evolution of oxygenic photosynthesis (Dismukes *et al.*, 2001; Kasting, 1993), the lack of

oxygen in an atmosphere would have resulted in lower overall biological activity (Des Marais, 2000). The capacity of oxygen to help provide abundant energy directly via aerobic respiration would simply not have existed. The production of a small amount of oxygen by photolysis of H₂O vapor in the upper atmosphere would have indirectly produced some oxidized inorganic electron acceptors, but nowhere near the quantity that have been available ever since the atmosphere was oxidized.

Today, because of the presence of oxygen in our atmosphere and dissolved oxygen in surface and near-surface ground water, aerobic microorganisms dominate the Earth's surface with anaerobic prokaryotes becoming increasingly important with depth below the surface. During the early- to mid-Archean, prior to the evolution of oxygenic photosynthetic bacteria, or presumably, on any geochemically reducing planetary body possessing an 'early stage' biosphere, anaerobic organisms would have been the dominant group (Kasting, 1993).

Below the aerobic zone (or beginning at the surface of an anaerobic planet), a progression of different anaerobes (e.g., metal reducers, sulfate reducers and methanogens) exists (would exist), most of which (from Earth's lessons) utilize organic matter for their metabolism (Lovley and Chapelle, 1995; Lovley and Klug, 1983). Under moderately reducing conditions, Fe- and Mn-reducing bacteria couple the oxidation of organic matter to the reductive dissolution of high-surface-area Fe- and Mn-oxyhydroxides (Lovley and Phillips, 1988; Myers and Nealson, 1988). Below the zone of metal reduction, sulfate-reducing bacteria (SRB) oxidize low-molecular-weight organic compounds, utilizing sulfate as the terminal electron acceptor, forming and releasing hydrogen sulfide as a byproduct of metabolism (Anderson *et al.*, 1998; Donald and Southam, 1999). Bacterial methane production occurs under even more reducing conditions than microbial sulfate reduction. Methanogenesis (reaction [5]) is also the dominant anaerobic prokaryotic process that cycles inorganic carbon back into the organic carbon pool (reaction [6]):



Biogenic methane can migrate out of the zone of methanogenesis and become an organic carbon

source (electron donor) to heterotrophic bacteria elsewhere. This redox zonation going from aerobic → Fe–Mn-oxide reduction → sulfate reduction → methanogenesis is a universal phenomenon on the present-day Earth, as demonstrated for example in marine sediment pore waters (e.g., Drever, 1997).

Since prokaryotes can only transport soluble compounds across their cell envelopes, they have evolved various metabolic ways in which to obtain nutrients from solid substrates, for example, Fe- or Mn-oxides, to support their growth. The ability to alter the chemistry of their surrounding environment on the scale of tens to hundreds of nanometers (Purcell, 1977), is what enables prokaryotes to promote the dissolution of otherwise 'stable' mineral phases (based on bulk fluid chemistry) or the precipitation of minerals under apparently undersaturated conditions. Remarkably, prokaryotes have recently demonstrated the capacity to produce electrically conductive nanowires under electron acceptor limiting conditions (Gorby *et al.*, 2006), thus demonstrating that biogeochemical cycling, which is not directly evident from fluid chemistry or mineralogy, may still be possible via interspecies electron transfer (Gorby *et al.*, 2006).

10.12.2.4 The Growth of Microorganisms

The small size (submicrometer to micrometer-scale) and resulting, high surface area:volume ratio of bacteria helps facilitate their growth. Since access to nutrients is based on diffusion and active uptake transporters at the cell surface, the greater the surface to volume ratio, the greater the capacity to obtain nutrients (Koch, 1996; Pirie, 1973; Purcell, 1977). Also, since individual prokaryotes only possess a mass of $\sim 1 \times 10^{-12}$ g (wet weight; Neidhardt *et al.*, 1990), extremely low amounts of dissolved nutrients can support their growth and survival. On Earth, prokaryotes are critical components of the food chain because they utilize dilute concentrations of inorganic and organic nutrients that are unavailable directly to higher, eukaryotic organisms (Fenchel and Jorgensen, 1977). The ability of prokaryotes to metabolize organic compounds in natural systems is dependent on the chemical's properties and the composition and mineralogy of the soil (Knaebel *et al.*, 1994). For example, Scow and Hutson (1992) found significant decreases in degradation of organic carbon in the presence of kaolinite, owing to the



Figure 4 A photograph of microbial mats growing near Atlin, BC, Canada. Prior to the evolutionary development of predators, approx. one half billion years ago, much of the photic biosphere on Earth presumably looked similar to this.



Figure 5 A photograph of an endolithic bacterial colonization (the blackening) of aeolian sandstone in potholes at Bartlett Wash, near Moab, NM (see [Chan et al. \(2005\)](#)).

effects of sorption and reduced diffusion from the clay surface.

Individual prokaryote species often grow as microcolonies ([Hall-Stoodley et al., 2004](#)), tens of microns in diameter, and can occur in association with other microcolonies within biofilms. Where freshwater–hypersaline aquatic conditions are stable for a long periods of time and where nutrients and favorable redox conditions exist for rapid growth, these ‘biofilms’ can reach large dimensions, up to centimeters or even meters in plan view and a millimeter or more in thickness, forming significant microbial mats on sediment/rock surfaces ([Figure 4](#); [Nealson and Stahl, 1997](#)). However, under arid environmental conditions where water activity is limited, for example, hot and cold desert environments ([Friedmann and Ocampo, 1976](#)), habitable microenvironments are found within soil or within rock ([Figure 5](#)), where these endolithic microbial communities display an intimate relationship with the substrate they inhabit. The extracellular polymeric substances (EPSs or capsule, [Figure 6](#); [Moser et al., 2003](#); [Wanger et al., 2006](#)) produced by these microorganisms serve many purposes, including protection of the colonies from the ‘external’ environment and providing a unique microenvironment or habitable zone in which the organisms can control the redox and pH conditions ([Roden et al., 2004](#)). In addition to linking individual microbial cells and colonies, EPS-rich biofilms can also play an important role in the stabilization of

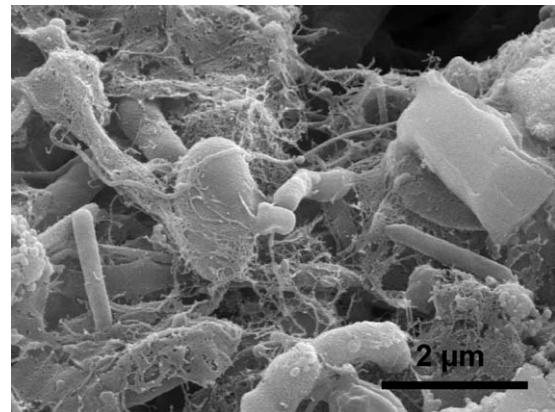


Figure 6 A scanning electron micrograph of a bacterial biofilm collected from a groundwater outflow into an oxygenated mining environment, 3.1 km below land surface in the Witwatersrand Basin, Republic of South Africa. Note the spider-web like appearance of the bacterial exopolymer material. Electron micrograph provided by M. Lengke.

sediment surfaces by cementing mineral particles together (see [Figures 6 and 2](#); [Stahl, 1994](#)), thus prolonging the stability of a particular habitat. While prokaryotes do not have direct control over the chemistry of the fluid phase that they encounter when attached to a mineral surface, attachment to a surface is advantageous in that it allows them to persist or hold their position within a presumably, favorable geochemical flow, letting the nutrients come to them ([Figure 7](#); [Emerson and Revsbech, 1994](#); [Purcell, 1977](#)).

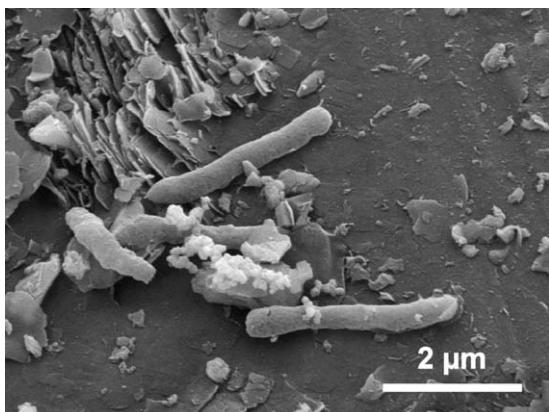


Figure 7 A scanning electron micrograph of a natural fracture face collected 2.8 km below land surface demonstrated that the deep subsurface possesses very sparse bacterial ‘biofilms’ (see Wanger *et al.* (2006)).

10.12.3 Geology and Life

10.12.3.1 Lithotrophy

Lithotrophic or ‘rock-eating’ prokaryotes are broadly characterized as organisms capable of utilizing inorganic compounds, for example, H_2S or minerals to generate energy, and can produce geologically significant phenomena. For instance, metal sulfide oxidation produces acidic environments such as acid mine drainage systems (Nordstrom and Southam, 1997; Singer and Stumm, 1970) and supergene copper deposits (Enders *et al.*, 2006).

Microorganisms are also active agents of chemical weathering of silicate minerals (e.g., Brantley *et al.*, 2001; Hiebert and Bennett, 1992; Rogers *et al.*, 1998; Figure 2) and if reactive organic matter is abundant, mineral dissolution can occur very rapidly. Silica weathering (both as grains and cement) is a general phenomenon resulting from microbial corrosion in order to concentrate elements limited in their environment, for example, iron and phosphate (Banerjee *et al.*, 2000), and micronutrients, such as Fe, Co, Zn, Mo, Cu, or Ni, which are represented in trace amounts in all terrestrial materials (Bennett *et al.*, 1996). Moreover, if apatite occurs in a silicate rock, it will be the first mineral to be dissolved out because phosphate is an essential ingredient for life, that is, a bio-limiting nutrient (Taunton *et al.*, 2000). An added advantage of silicate weathering is the creation of suitable habitats for the growth of endolithic microbial communities in water-limited ecosystems (Chan *et al.*, 2005; Friedmann 1980, 1982). This process also results in the exfoliation of silicate rocks that is

typical of hot or cold desert environments, whereby the metabolic activity of microbes inhabiting the protective microhabitat beneath the exposed surface of the rocks weakens the rock in layers parallel to its surface (Parnell *et al.*, 2004).

The importance of mineral weathering in supporting the biosphere has led to the general hypothesis that life on Earth or any other rocky planetary body would ultimately be limited without some form of plate tectonics or lithosphere renewal providing nutrients from the more reduced lithosphere below. However, once established, subsurface colonization of planetary bodies would extend life for much longer periods of time as the planet cools, allowing the deeper penetration of life into the lithosphere.

10.12.3.2 Microorganisms and Mineral Formation

Given the right conditions, that is, sufficient mineral ions in the aqueous medium, microorganisms and their biofilms or mats can be rapidly fossilized (Figure 1; Schultz-Lam *et al.*, 1995; Westall *et al.*, 1995). Fossilization takes place either (1) by the complexation of mineral ions to functional groups in the organic material, gradually replacing the organic structure and, at the same time, trapping the degrading organic molecules in the polymerizing mineral matrix, or (2) when amorphous minerals formed in the aqueous phase bind to the cell envelope (Konhauser *et al.*, 2003; Rancourt *et al.*, 2005). The organic matter in the fossilized microorganisms is typically degraded but parts of it may remain trapped in the mineral matrix. Alternatively, in modern, oxidizing hydrothermal environments, the organic matter can be completely oxidized, although the microbial cast still remains (Konhauser, 1988). It is hard to imagine that any biosphere could exist without becoming at least partially fossilized. Even the prebiotic realm would be equally subject to ‘fossilization’, since the organic macromolecules making it up will also consist of functional groups that can chelate mineral ions (Southam and Donald, 1999).

The formation of many secondary minerals in natural as well as laboratory systems is catalyzed by microorganisms (Figure 8; see Lowenstam (1981)). Metal sorption and mineral nucleation onto microbial cell envelopes are two different, passive mineralization processes. Metal sorption occurs via ion-exchange reactions through competition between hydronium ions, alkaline earth ions, and

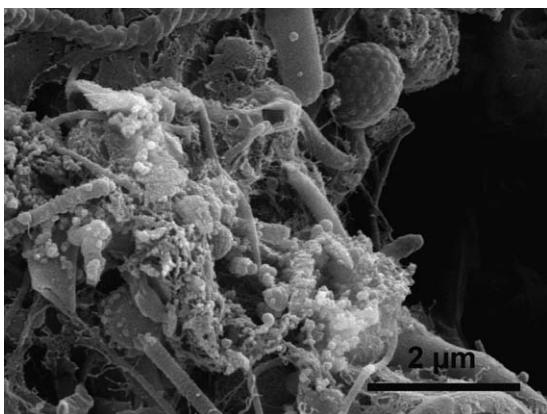
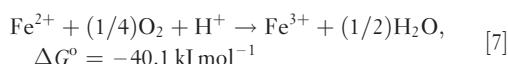


Figure 8 A scanning electron micrograph of a biofilm collected from a borehole flowing from the deep subsurface, 3.1 km below land surface in the Witwatersrand Basin, Republic of South Africa. Note the fine-grained, amorphous iron silicate minerals that have precipitated within the biofilm. Electron micrograph provided by M. Lengke.

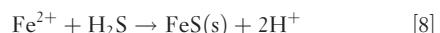
heavy metals for anionic reactive sites on microbial surfaces (Beveridge, 1981; Urrutia and Beveridge, 1994). Many cell envelope polymers have a demonstrated capacity to bind metallic ions and promote mineral nucleation (Beveridge, 1981; Fein *et al.*, 1997).

Active surface-mediated mineralization occurs either by the direct transformation of metals into unstable forms (Lovley *et al.*, 1991), or by the formation of metal reactive by products (Fortin *et al.*, 1994). For example, cyanobacteria have the capacity to precipitate carbonates (Thompson and Ferris, 1990; Thompson *et al.*, 1997) via the creation of alkaline conditions immediately around the cell from photosynthetic activity (Figure 4). However, the formation of unstable forms or metal reactive byproducts depends on the geochemical environment in which they are formed. One hypothesis for the deposition of banded iron formations in the Late Archaean is that iron oxide precipitation was mediated by the metabolic activity of oxygenic photosynthetic cyanobacteria. Today and since the oxygenation of Earth's atmosphere, the oxidation of iron (reaction [7]) occurs within the aerobic/anaerobic interface of the Earth surface where the reduced iron encounters oxidizing conditions



(Ehrlich, 1975; Edwards, *et al.*, 2004; Emerson and Revsbech, 1994; Little *et al.*, 2004). Dissimilatory

metal reducing prokaryotes are best known for their ability to utilize minerals for their electron acceptors resulting in the solubilization of Fe (Lovley and Philips, 1988) and Mn (Myers and Nealson, 1988) oxides and any co-precipitated base metals, as well as the formation of magnetite, siderite, chromium hydroxide, and uranite (Ishibashi *et al.*, 1990; Lovley *et al.*, 1991; Pedersen and Ekendahl, 1990). The most common metal sulfide attributed to biogenic activity is iron sulfide (Tuttle *et al.*, 1969):



These dissimilatory processes combined with bacterial surface catalysis are responsible for cell surface mineral formation in these systems.

Although individual bacteria are extremely small, bacteria have played a role in the formation of geologically significant deposits of carbonate, banded iron formations, and phosphorites (Konhauser *et al.*, 2002; Riding, 2000; Youssef, 1965; see Figure 9). From a

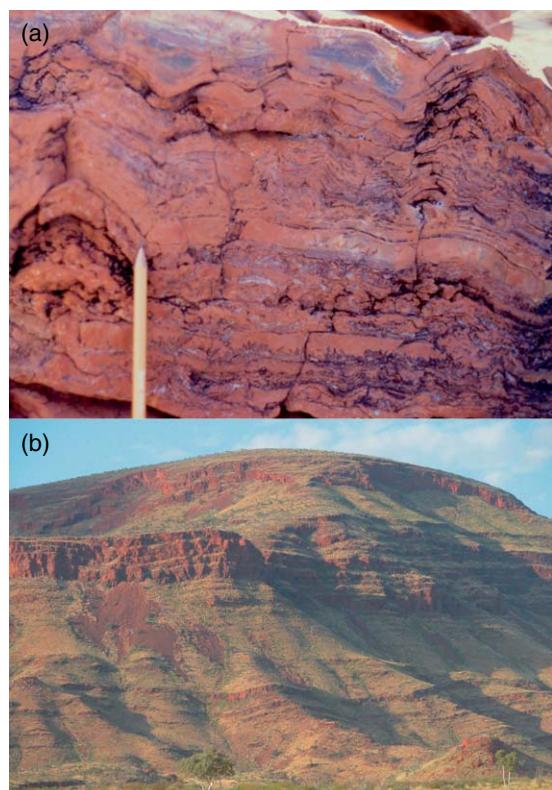


Figure 9 (a) Small, conical stromatolites from the Early Archean (>3.46 billion years old) North Pole Dome of the Pilbara (photograph D. Marchesini). (b) A photograph of the Brockman Formation, a Banded Iron Formation found in the Hamersley Range, Pilbara, Western Australia.

geochemical perspective, the contribution of bacteria to the formation of ancient deposits is often inferred by the enrichment of stable light isotopes by microbial mineral-forming processes (Kakegawa, 2001; Schidlowski, 1988). Mineral signatures then can serve as biomarkers when they are associated with fossils (Cady *et al.*, 2003; Rasmussen, 2000; Southam *et al.*, 2001; Tazaki *et al.*, 1992), when they possess isotopic signatures that point to biochemical (enzymatic) processing of dissolved solutes, or when realistic, abiotic geological phenomena cannot explain their occurrence (see review in Westall and Southam (2006)).

10.12.3.3 Habitable Conditions on Early Earth (and Other Early Rocky Planets)

As we saw briefly in Section 10.12.1, the early Earth would have presented a very different habitat compared with the modern planet (see reviews in Nisbet and Fowler (2004) and Nisbet and Sleep (2001)). The hotter mantle meant that there was a greater amount of volcanic and hydrothermal activity (perhaps similar to that observed in the central portion of Figure 1), as well as faster recycling of the crust (Arndt, 1994). The atmosphere was composed of mostly CO₂, together with CO, NH₃, N₂ and probably a certain amount of CH₄ (the latter possibly at least partly biological; Pavlov *et al.*, 2001; Rye *et al.*, 1995). Oxygen levels were low (<0.2% of present atmospheric levels; Kasting, 1993), the oxygen being produced by photolysis of water vapor in the upper atmosphere. Based on the high CO₂ concentrations in the atmosphere (Rye *et al.*, 1995), the oceans were probably slightly acidic (Grotzinger and Kasting, 1993). These acidic conditions, combined with the rapid cycling of hydrothermal fluids (Duchać and Hanor, 1987; Paris *et al.*, 1985), would have resulted in extensive mineral weathering. Higher salinity (de Ronde *et al.*, 1997) and higher temperatures (50–80°C according to Knauth and Lowe (2003)) were also characteristic of the Early-Mid Archean oceans. The lack of significant oxygen in the atmosphere meant that UV radiation levels were high, that is, up to 1000 times present levels in a worst case scenario (Cockell and Raven, 2004; a habitat that could only be considered extreme by our standards). However, the deleterious effects of UV radiation on the biosphere may have been mitigated by large amounts of dust and aerosols in the atmosphere from volcanic eruptions, meteoritic impacts (Lowe *et al.*, 2003), and possibly water vapor (Lammer *et al.*, 2005) or CH₄ smog (Lovelock, 1988). Also, the

coexistence of reducing and oxidized gases in the prebiotic atmosphere, combined with the input of extraterrestrial organic carbon (e.g., Brown *et al.*, 2000), would have provided abundant untapped chemical energy. At some point, life must have evolved to exploit this chemical combination (Catling and Claire, 2005), for example, methanogens, that utilize the redox disequilibrium by combining H₂ and CO₂ to make methane (reaction [4]). Isotopic studies of fluid inclusions in Archean rocks have found that the composition of the oceans has remained somewhat constant over time (Holland, 1984; Kasting *et al.*, 1993); the maintenance of this reducing to microaerophilic environment for 1–2 billion years (Kasting *et al.*, 1993) would be considered ideal for a life form, which originated in it.

10.12.3.4 The Earliest Traces of Life

We would describe the early life forms on Earth as being anaerobic, halotolerant, UV resistant (see Daly *et al.*, 1994), and thermophilic (see Stetter *et al.*, 1990). What is intriguing about the remnants of the oldest life on Earth is (1) the modern-looking aspect of the individual microorganisms, their colonies, growth as biofilms and mats (Figure 10), (2) their diversity, and (3) their wide environmental distribution (see review in Westall and Southam (2006)). While early stromatolitic constructions (Allwood *et al.*, 2006; Hofman *et al.*, 1999) are orders of magnitude smaller (Figure 9(a)) than those of the Late Archean/Proterozoic epochs (Figure 9(b)) (Awramik and Sprinkle, 1999), this is not the record of primitive life forms (Westall and Southam, 2006). While ‘organic-rich’ shales deposited in deeper waters during this period testify to a relatively abundant input of carbonaceous material from shallower water environments (Walsh and Lowe, 1999; Tice and Lowe, 2004) or from the planktonic photic zone (Walsh and Lowe, 1999), the main limitation to productivity appears to be the fact that oxygenic photosynthesis and the corresponding aerobic heterotrophs had not yet developed (Des Marais, 2000).

The advanced level of evolution demonstrated by the remnants of life at 3.5 Ga has implications for the timing of the origin of life. Did the predicted planet-sterilizing impacts (Sleep *et al.*, 1989; Nisbet and Sleep, 2001; Ryder, 2002) during the late heavy bombardment between 4.0–3.85 Ga (Maher and Stevenson, 1988) completely wipe out all traces of life, if it had evolved before, or was life able to survive in protected subsurface ecosystems? Note that an

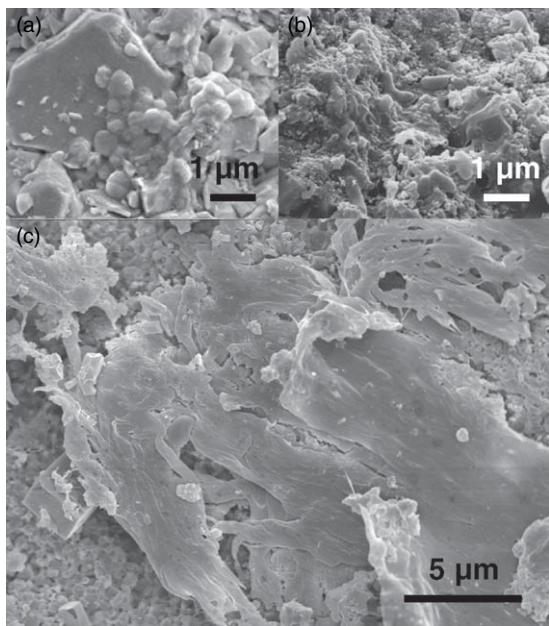


Figure 10 Fossilized Early Archean microorganisms (3.47–3.34 billion years): (a) small colony of small coccoid microfossils (arrow) that colonized a marine, subsurface sedimentary habitat (see [Westall et al. \(2006a\)](#)); (b) small colony of rod-shaped to vibrioid shaped microfossils (arrow) that were closely associated anaerobic, photosynthetic filaments forming a well-developed microbial mat (c) on beach sediments in an evaporitic environment (see [Westall et al. \(2006b\)](#)).

alternative hypothesis proposes that only the topmost 400 m of the oceans would have been evaporated during the worst of the impact events (Ryder, 2002). The occurrence of hyperthermophilic microorganisms in the lowest branches of the evolutionary tree of life (Pace, 1997) is generally considered to imply that early life must have passed through severe (catastrophic) environmental conditions, such as an (almost) planet-sterilizing impact, and that life forms in protected habitats, such as deep-sea hydrothermal vents, survived, whereas life forms in the more exposed habitats did not. Whatever the timing necessary for life to appear ([Lazcano and Miller, 1996](#), suggest 10 My), a rocky planet with a hydrosphere can support prokaryote life forms.

10.12.4 Conclusions for Geology, Life and Habitability beyond the Earth

The ingredients necessary for life to originate (liquid water in contact with rocks/minerals, abundant

organic molecules (exogenous or endogenous), elements such as hydrogen, oxygen, nitrogen, and sulfur, and energy sources (geochemical, sunlight)) were abundant on the early Earth. They were also abundant elsewhere in the early solar system (see Chapter 10.11) and presumably in the universe. It is therefore legitimate to surmise that life could have arisen independently wherever these initial conditions are stable for the length of time that is necessary for the process say about 10–100 My (cf. [Lazcano and Miller, 1996](#)). Once started, from a microbiological point of view, primitive life forms can exist wherever there are the essential nutrients and potential redox reactions that could support growth. Apart from the classical terrestrial planets in the solar system (Venus and Mars), Europa has also recently excited interest because of the potential for liquid water in contact with a rocky core ([Bhattacherjee and Chela-Flores, 2004](#); [Greenberg, 2005](#)) and the presence of water ice on Encelade is equally stimulating ([Brown et al., 2006](#); see Chapter 10.15).

The possibility of extraterrestrial life on Mars has fascinated humans for over a hundred years: Percival Lowell in 1910 started the ball rolling with an account of Martian life based on Schiaparelli's (1879) drawing of what he thought were channels (or intersecting straight lines) on the surface of Mars. Although various missions over the last 30 years have discounted these fanciful theories, it is clear that, early in its history, there was a significant inventory of water on the surface of the planet ([Jaumann et al., 2002](#); [Bibring et al., 2006](#)), albeit probably not sufficient for an ocean covering the Northern Hemisphere ([Parker et al., 1993](#)). Notably, Mars possessed an active hydrosphere during the time when life originated on Earth. As we have seen above, microbial life does not require oceans of water in order to function: once it has got going, it just requires continual access to a very small amount of carbon, nutrients, and energy. Given the access to the basic requirements for an origin of life (liquid water, organics, and rocks/minerals), it is possible that life did arise independently on the red planet. Therefore, Mars may contain a record of fossilized primitive microorganisms or their associated mineralogical biomarkers, similar to those that are found on Earth. These traces of early Martian life will likely occur in the rocks of the Southern Hemisphere, which are largely Noachian in age and have apparently undergone aqueous alteration ([Bibring et al., 2006](#)). Of course, boulders or meteorites of Southern Hemisphere material could be found elsewhere on

Mars, or even on the moons of Mars, Phobos and Demos, since Mars, together with the other inner solar system planets, experienced a period of late bombardment between 4.0 and 3.8 billion years ago. It is equally possible that the Noachian age rocks will contain traces of the prebiotic stage of life and the appearance of the first cells, that is, the critical period of life on Earth's history that is missing on Earth – if we are capable of recognizing these traces.

If life established itself on Mars during the Noachian age, could it still be present today on Mars and, if so, where? The loss of liquid water on Mars' surface and the cessation of any sort of plate tectonic activity (if there ever was any) due to the rapid cooling of the planet, limits one of the important criteria for life, that is, a flux of reduced or energetic redox species. As we noted above, micro-organisms can survive a wide range of extreme conditions but there are certain combinations that are lethal and it appears that the combined low water activity, high UV radiation plus the presence of some kind of oxidant species that oxidizes labile organic components (i.e., those that are necessary for life), make the surface of the planet hostile to any long-term survival. However, Mars is not completely cold, and ongoing volcanic activity (Neukum *et al.*, 2004) indicates that there could be hydrothermal activity in its subsurface, as indicated by occasional surface exposure (Malin and Edgett, 2000). The seepage of methane that has recently been detected (Formisano *et al.*, 2004) may also indicate hydrothermal melting of CH₄ clathrates (Prieto-Ballesteros *et al.*, 2006), although some argue for a biological origin of the methane, for example, Tung *et al.* (2005). This activity would provide the necessary criteria/conditions for life, similar to that provided to bacteria in Earth's subsurface (Boston *et al.*, 1992; Wanger *et al.*, 2006). Thus, Mars could still host life in its subsurface.

While we do not know the exact conditions necessary to create life, the basic requirements of a habitable extraterrestrial body (assuming Earth is a representative sample) are fairly well understood. Habitable locations in our solar system today, that is, places where life can be sustained, are not as restrictive as those where life could originate, and include Mars' subsurface (Boston *et al.*, 1992), as well as the ocean on Europa (Chyba and Phillips, 2001, see Chapter 10.15). It is also intriguing to realize that, in the future as the Sun gets hotter and comes into the red giant phase of its history, these planets and possibly others in the outer solar system will be able to

host life, if not see it appear independently. Thus, the possibility of finding traces of extinct or extant life on other planets in the Solar System is strong and the future for life in the Solar System is optimistic, even though the habitability of our own planet will slowly decrease starting in about 1.5 Ga (Franck *et al.*, 2006).

References

- Allwood AC, Walter MR, Kamber BS, Marshall CP, and Burch IW (2006) Stromatolite reef from the Early Archaean era of Australia. *Nature* 441: 714–718.
- Anderson IK, Ashton JH, Boyce AJ, Fallick AE, and Russell MJ (1998) Ore depositional processes in the Navan Zn-Pb deposit, Ireland. *Economic Geology* 93: 535–563.
- Arndt NT (1994) Archean komatiites. In: Condie KC (ed.) *Archean Crustal Evolution*, pp. 11–44. Amsterdam: Elsevier.
- Atlas RM and Bartha R (1997) *Microbial Ecology: Fundamentals and Applications* 4th edn. 306 pp. Menlo Park, CA: Benjamin Cummings.
- Awrakim SM and Sprinkle J (1999) Proterozoic stromatolites: The first marine evolutionary biota. *Historical Biology* 13: 241–253.
- Banerjee M, Whitton BA, and Wynn-Williams DD (2000) Phosphatase activities of endolithic communities in rocks of the Antarctic dry valleys. *Microbial Ecology* 39: 80–91.
- Barbier B and Brack A (2002) Conformation-controlled hydrolysis of polyribonucleotides by sequential basic polypeptides. *Journal of American Chemical Society* 114: 3511–3515.
- Barns SM and Nierzwicki-Bauer SA (1997) Microbial diversity in ocean, surface and subsurface environments. In: Banfield JF and Nealson KH (eds.) *Reviews in Mineralogy 35: Geomicrobiology: Interactions between Microbes and Minerals*, pp. 35–79. Chantilly, VA: Mineralogical Society of America.
- Bechtel A, Sieh Y-N, Pervaz M, and Puttmann W (1996) Biodegradation of hydrocarbons and biogeochemical sulfur cycling in the salt dome environment: Inferences from sulfur isotope and organic geochemical investigations of the Bahloul formation at the Bou Grine Zn/Pb ore deposit (Tunisia). *Geochimica Cosmochimica Acta* 60: 2833–2855.
- Bennett PC, Hieber FK, and Choi WJ (1996) Rates of microbial weathering of silicates in ground water. *Chemical Geology* 132: 45–53.
- Beveridge TJ (1981) Ultrastructure, chemistry, and function of the bacterial wall. *International Review of Cytology* 72: 229–317.
- Bhattacharjee AB and Chela-Flores J (2004) Search for bacterial waste as a possible signature of life on Europa. In: Seckbach J, Chela-Flores J, Owen T, and Raulin F (eds.) *Life in the Universe, Cellular Origin and Life in Extreme Habitats and Astrobiology*, vol.7, pp. 257–260. Dordrecht, The Netherlands: Springer.
- Bibring J-P, Langevin Y, Mustard J-F, *et al.* (2006) Global mineralogical and aqueous Mars history derived from OMEGA/Mars Express Data. *Science* 312: 400–403.
- Boston PJ, Ivanov MV, and McKay CP (1992) On the possibility of chemosynthetic ecosystems in subsurface habitats on Mars. *Icarus* 95: 300–308.
- Brack A (2004) The chemistry of life's origins. In: Seckback J (ed.) *Life in the Universe*, pp. 59–73. Dordrecht: Kluwer.
- Brack A (2005) L'exobiologie ou l'origine chimique de la vie. In: Halbwachs JL, Egret D, and Hameury JM (eds.)

- Formation planétaire et exoplanètes. Ecole CNRS de Goutelas XXVIII.
- Brantley SL, Lierman L, Bau M, and Wu S (2001) Uptake of trace metals and rare elements from hornblende by a soil bacterium. *Geomicrobiology Journal* 18: 37–61.
- Brill WJ (1975) Regulation and genetics of bacterial nitrogen fixation. *Annual Review of Microbiology* 29: 109–129.
- Brock JJ, Logan GA, Buick R, and Summons RE (1999) Archean molecular fossils and the early rise of eukaryotes. *Science* 285: 1033–1036.
- Brown PG, Hildebrand AR, Zolensky ME, et al. (2000) The fall, orbit, and composition of the Tagish Lake Meteorite: A new type of carbonaceous chondrite. *Science* 290: 320–325.
- Brown RH, Clark RN, Buratti BJ, et al. (2006) Composition and physical properties of Enceladus' surface. *Science* 311: 1425–1428.
- Cady SL, Farmer JD, Grotzinger JP, Schopf WJ, and Steele A (2003) Morphological biosignatures and the search for life on Mars. *Astrobiology* 3: 351–368.
- Catling DC and Claire MW (2005) How Earth's atmosphere evolved to an oxic state: A status report. *Earth and Planetary Science Letters* 237: 1–20.
- Cavanaugh CM, Gardiner SL, Jones ML, Jannasch HW, and Waterbury JB (1981) Prokaryotic cells in the hydrothermal vent tube worm *Riftia pachyptila*: Possible chemoaautotrophic symbionts. *Science* 213: 340–342.
- Chan M, Moser K, Davis JM, Southam G, Hughes K, and Graham T (2005) Desert potholes: Life in the fast pool. *Aquatic Geochemistry* 11: 279–302.
- Chapelle FH (1993) *Ground-Water Microbiology and Geochemistry*, 448 pp. New York: John Wiley.
- Chyba CF and Phillips CB (2001) Possible ecosystems and the search for life on Europa. *PNAS* 98: 801–804.
- Cockell CS and Raven JA (2004) Zones of photosynthetic potential on Mars and the early Earth. *Icarus* 169: 300–310.
- Colwell FS, Onstott TC, Dilwiche ME, et al. (1997) Microorganisms from deep, high temperature sandstones: Constraints on microbial colonization. *FEMS Microbiology Reviews* 20: 425–435.
- Daly MJ, Ouyang L, and Minton KW (1994) *In vitro* damage and recA-dependent repair of plasmid and chromosomal DNA in the radioresistant bacterium *Deinococcus radiodurans*. *Journal of Bacteriology* 176: 3508–3517.
- Deamer DW (1998) Membrane compartments in prebiotic evolution. In: Brack A (ed.) *The Molecular Origins of Life: Assembling Pieces of the Puzzle*, pp. 189–205. Cambridge: Cambridge University Press.
- Des Marais DJ (2000) When did photosynthesis emerge on Earth?. *Science* 289: 1703–1705.
- de Ronde CEJ, Channer DMd, Faure K, Bray CJ, and Spooner ETC (1997) Fluid chemistry of Archean seafloor hydrothermal vents: implications for the composition of circa 3.2 Ga seawater. *Geochimica Cosmochimica Acta* 61: 4025–4042.
- Dismukes GC, Klimov VV, Baranov SV, Kozlov YN, DasGupta J, and Tyrshkin A (2001) The origin of atmospheric oxygen on Earth: The innovation of oxygenic photosynthesis. *PNAS* 98: 2170–2175.
- Donald R and Southam G (1999) Low temperature anaerobic bacterial diagenesis to ferrous monosulfide to pyrite. *Geochimica et Cosmochimica Acta* 63: 2019–2023.
- Drever JI (1997) *The Geochemistry of Natural Waters* 3rd edn. 436 pp. Upper Saddle River, NJ: Prentice Hall.
- Duchac KC and Hanor JS (1987) Origin and timing of the metasomatic silification of an Early Archaean komatiite sequence, Barberton Mountain Land, South Africa. *Precambrian Research* 37: 125–146.
- Edwards KJ, Bach W, McCollom TM, and Rogers DR (2004) Neutrophilic iron-oxidizing bacteria in the ocean: Their habitats, diversity, and roles in mineral deposition, rock alteration, and biomass production in the deep-sea. *Geomicrobiology Journal* 21: 393–404.
- Ehrlich HL (1975) The formation of ores in the sedimentary environment of the deep sea with microbial participation: The case for ferromanganese concretions. *Science* 191: 36–41.
- Emerson D and Revsbech NP (1994) Investigation of an iron-oxidizing microbial mat community located near Aarhus, Denmark: Field studies. *Applied and Environmental Microbiology* 60: 4022–4031.
- Enders MS, Southam G, Knickerbocker C, and Titley SR (2006) The role of microorganisms in the supergene environment of the Morenci porphyry copper deposit, Greenlee County, Arizona. *Economic Geology* 101: 59–70.
- Fenchel TM and Jorgensen BB (1977) Detritus food chains of aquatic ecosystems: The role of bacteria. *Advances in Microbial Ecology* 1: 1–58.
- Fein JB, Daughney CJ, Yee N, and Davis T (1997) A chemical equilibrium model of metal adsorption onto bacterial surfaces. *Geochimica Cosmochimica Acta* 61: 3319–3328.
- Ferris JP, Hill AR, Lui R, and Orgel LE (1996) Synthesis of long prebiotic oligomers on mineral surfaces. *Nature* 381: 59–61.
- Formisano V, Atreya S, Encrenaz T, Ignatiev N, and Giuranna M (2004) Detection of methane in the atmosphere of Mars. *Science* 306: 1758–1761.
- Fortin D, Southam G, and Beveridge TJ (1994) An examination of iron sulfide, iron–nickel sulfide and nickel sulfide precipitation by a *Desulfotomaculum* species and its nickel resistance mechanisms. *FEMS Microbiology Ecology* 14: 121–132.
- Fox SW and Dose K (eds.) (1977) *Molecular Evolution and the Origin of Life*. New York: Marcel Dekker.
- Franck S, Bounama C, and von Bloh W (2006) Causes and timing of future biosphere extinctions. *Biogeosciences* 3: 85–92.
- Friedmann EI (1980) Endolithic microbial life in hot and cold deserts. *Origin of Life* 10: 223–235.
- Friedmann EI (1982) Endolithic microorganisms in the Antarctic cold desert. *Science* 215: 1045–1052.
- Friedmann EI and Ocampo R (1976) Endolithic blue-green algae in the dry valleys: Primary producers in the Antarctic desert ecosystems. *Science* 193: 1247–1249.
- Furnes H, Banerjee NR, Muehlenbachs K, Staudigel H, and de Wit M (2004) Early life recorded in Archean pillow lavas. *Science* 304: 578–581.
- Goldman M and Wilson DA (1977) Growth of *Sporosarcina ureae* in defined media. *FEMS Microbiology Letters* 2: 113–115.
- Goodwin S and Zeikus JG (1987) Physiological adaptations of anaerobic bacteria to low pH: metabolic control of proton motive force in *Sarcina ventriculi*. *Journal of Bacteriology* 169: 2150–2157.
- Gorbay YA, Yanina S, McLean JS, et al. (2006) Electrically conductive bacterial nanowires produced by *Shewanella oneidensis* strain MR-1 and other microorganisms. *PNAS* 103: 11358–11363.
- Greenberg R (2005) Europa – The Ocean Moon: Search for an Alien Biosphere, 380 pp. Chichester, UK: Springer Praxis Publishing.
- Grotzinger JP and Kasting JF (1993) New constraints on Precambrian ocean composition. *Journal of Geology* 101: 235–243.
- Hall-Stoodley L, Costerton JW, and Stoodley P (2004) Bacterial biofilms: From the environment to infectious disease. *Nature Reviews Microbiology* 2: 95–108.
- Heising S and Schink B (1998) Phototrophic oxidation of ferrous iron by a *Rhodomicrobium vannielli* strain. *Microbiology* 144: 2263–2269.

- Hiebert F and Bennett PC (1992) Microbial control of silicate weathering in organic-rich ground water. *Science* 258: 278–281.
- Hofmann HJ, Grey K, Hickman AH, and Thorpe RI (1999) Origin of 3.45 Ga coniform stromatolites in Warrawoona Group, Western Australia. *Geological Society of America Bulletin* 111: 1256–1262.
- Holland HD (1984) The chemical evolution of the atmosphere and oceans. Princeton: Princeton University Press.
- Holm NG and Charlot J-L (2001) Initial indications of abiotic formation of in the Rainbow ultramafic hydrothermal system, Mid-Atlantic Ridge. *Earth and Planetary Science* 191: 1–8.
- Inoue T and Orgel LE (1982) Oligomerization of (guanosine 5'-phospho)-2-methylimidazolide on poly(c). An RNA polymerase model. *Journal of Molecular Biology* 162: 201–217.
- Ishibashi Y, Cervantes C, and Silver S (1990) Chromium reduction in *Pseudomonas putida*. *Applied and Environmental Microbiology* 56: 2268–2270.
- Jaumann R, Hauber E, Lanz J, Hoffmann H, and Neukum G (2002) Geomorphological record of water-related erosion on Mars. In: Horneck G and Baumstark-Kahn C (eds.) *Astrobiology*, pp. 89–109. Berlin: Springer; New York: Heidelberg.
- Kakegawa T (2001) Isotopic signatures of early life in the Archean oceans: Influence from submarine hydrothermal activities. In: Nakashima S, Maruyama S, Brack A, and Windley BF (eds.) *Geochemistry and the Origin of Life*, pp. 237–249. Tokyo: Universal Academic Press.
- Kappler A and Newman DK (2004) Formation of Fe(III)-minerals by Fe(II)-oxidizing photoautotrophic bacteria. *Geochimica Cosmochimica Acta* 68: 1217–1226.
- Kargel J (2004) Mars: A Warmer, Wetter Planet, 557 pp. Berlin: Springer.
- Kashefi K and Lovley DR (2003) Extending the upper temperature limit for life. *Science* 301: 934.
- Kasting JF (1993) Earth's early atmosphere. *Science* 259: 920–926.
- Kasting JF, Whitmire DP, and Reynolds RT (1993) Habitable zones around main sequence stars. *Icarus* 101: 108–128.
- Knaebel DB, Federle TW, McAvoy DC, and Vestal JR (1994) Effect of mineral and organic soil constituents on microbial mineralization of organic compounds in a natural soil. *Applied and Environmental Microbiology* 60: 4500–4508.
- Knauth LP and Lowe DR (2003) High Archean climatic temperature inferred from oxygen isotope geochemistry of cherts in the 3.5 Ga Swaziland Supergroup, South Africa. *Geological Society of America Bulletin* 115: 566–580.
- Koch AL (1996) What size should a bacterium be? A question of scale. *Annual Review of Microbiology* 50: 317–348.
- Konhauser KO, Hamade T, Morris RC, et al. (2002) Could bacteria have formed the Precambrian banded iron formations?. *Geology* 30: 1079–1082.
- Konhauser KO, Jones B, Reysenbach A-L, and Renaut RW (2003) Hot spring sinters: Keys to understanding Earth's earliest life forms. *Canadian Journal of Earth Sciences* 40: 1713–1724.
- Konhauser K (1998) Diversity of bacterial iron mineralization. *Earth-Science Reviews* 43: 91–121.
- Kunte HJ (2006) Osmoregulation in bacteria: Compatible solute accumulation and osmosensing. *Environmental Chemistry* 3: 94–99.
- Lammer H, Yu N, Kulikov T, et al. (2005) Stellar-planetary relations: Atmospheric stability as a prerequisite for planetary habitability. In: Dvorak R and Ferraz-Mello S (eds.) *A Comparison of the Dynamical Evolution of Planetary Systems. Proceedings of the 6th Alexander von Humboldt Colloquium* pp. 273–285. Kluwer.
- Lazcano A and Miller SL (1996) The origin and early evolution of life: Prebiotic chemistry, the pre-RNA world, and time. *Cell* 85: 793–798.
- Little CTS, Glynn SEJ, and Mills RA (2004) Four-hundred-and-ninety-million-year record of bacteriogenic iron oxide precipitation at sea-floor hydrothermal vents. *Geomicrobiology Journal* 21: 415–429.
- Ljungdahl LG and Eriksson K-E (1985) Ecology of microbial cellulose degradation. *Advances in Microbial Ecology* 8: 237–299.
- Lovley DR and Chapelle FH (1995) Deep subsurface microbial processes. *Reviews of Geophysics* 33: 365–381.
- Lovley DR and Klug MJ (1983) Sulfate reducers can outcompete methanogens at freshwater sulfate concentrations. *Applied and Environmental Microbiology* 45: 187–192.
- Lovley DR and Phillips EJP (1988) Novel mode of microbial energy metabolism: Organic carbon oxidation coupled to dissimilatory reduction of iron or manganese. *Applied and Environmental Microbiology* 54: 1472–1480.
- Lovley DR, Phillips EJP, Gorby YA, and Landa ER (1991) Microbial reduction of uranium. *Nature* 350: 413–416.
- Lovelock JE (1988) *The Ages of Gaia*. New York: W.W. Norton.
- Lowe DR (1980) Stromatolites 3,400-Myr old from the Archean of Western Australia. *Nature* 284: 441–443.
- Lowe DR and Byerly GR (1986) Early Archean silicate spherules of probable impact origin, South Africa and Western Australia. *Geology* 14: 83–86.
- Lowe DR, Byerly GR, Kyte FT, Shuklyukov A, Asaro F, and Krull A (2003) Spherule beds 3.47–3.24 billion years old in the Barberton Greenstone Belt, South Africa: A record of large meteorite impacts and their influence on early crustal and biological evolution. *Astrobiology* 3: 7–48.
- Lowell P (1910) *Mars as the Abode of Life*, 288 pp. New York: Macmillan.
- Lowenstam HA (1981) Minerals formed by organisms. *Nature* 211: 1126–1131.
- Maher KA and Stevenson DJ (1988) Impact frustration of the origin of life. *Nature* 331: 612–614.
- Malin MC and Edgett KS (2000) Evidence for recent groundwater seepage and surface runoff on Mars. *Science* 288: 2330–2335.
- Marshall KC (1988) Adhesion and growth of bacteria at surfaces in oligotrophic habitats. *Canadian Journal of Microbiology* 34: 503–506.
- Maurette M, Duprat J, Engrand C, et al. (2000) Accretion of neon, organics, CO₂, nitrogen and water from large interplanetary dust particles on the early Earth. *Planetary and Space Science* 48: 1117–1137.
- Miller SL (1953) A production of amino acids under possible primitive earth condition. *Science* 117: 528–529.
- Moser DP, Onstott TC, Fredrickson JK, et al. (2003) Evolution of microbial community structure and geochemistry in an ultradepth South African gold mine borehole. *Geomicrobiology Journal* 20: 517–548.
- Myers CP and Nealson KN (1988) Bacterial manganese reduction and growth with manganese oxide as the sole electron acceptor. *Science* 240: 1319–1321.
- Navarro-Gonzalez R, Rainey FA, Molina P, et al. (2003) Mars-like soils in the Atacama Desert, Chile, and the dry limit of microbial life. *Science* 302: 1018–1021.
- Nealson KH and Stahl DA (1997) Microorganisms and biogeochemical cycles: What can we learn from layered microbial communities? In: Banfield JF and Nealson KH (eds.) *Reviews in Mineralogy*, vol. 35: *Geomicrobiology: Interactions between Microbes and Minerals*, pp. 5–34. Chantilly, VA: Mineralogical Society of America.
- Neidhardt FC, Ingraham JL, and Schaechter M (1990) *Physiology of the Bacterial Cell – A Molecular Approach*, 506 pp. Sunderland, MA: Sinauer Associates, Inc.

- Neukum G, Jaumann R, Hoffmann H, et al. (2004) Mars: Recent and episodic volcanic, hydrothermal, and glacial activity revealed by the Mars Express High Resolution Stereo Camera (HRSC) experiment. *Nature* 432: 971–979.
- Newman DK and Banfield JF (2002) Geomicrobiology: How molecular-scale interactions underpin biogeochemical systems. *Science* 296: 1071–1077.
- Newman MJ and Rood RT (1977) Implications of solar evolution for the Earth's early atmosphere. *Science* 198: 1035–1037.
- Nisbet EG and Fowler CMR (2004) The early history of life. In: Holland HD and Turekian KK (eds.) *Treatise on Geochemistry, Vol. 8: Biogeochemistry*, pp. 1–39. Oxford: Elsevier-Pergamon.
- Nisbet EG and Sleep NH (2001) The habitat and nature of early life. *Nature* 409: 1083–1091.
- Nordstrom DK and Southam G (1997) Geomicrobiology of sulfide mineral oxidation. *Reviews in Mineralogy* 35: 361–390.
- Onstott TC, Lin L-H, Davidson M, et al. (2006) The origin and age of biogeochemical trends in deep fracture water of the Witwatersrand Basin, South Africa. *Geomicrobiology Journal* 23: 369–414.
- Pace NR (1997) A molecular view of microbial diversity and the biosphere. *Science* 276: 734–740.
- Paris I, Stanistreet IG, and Hughes MJ (1985) Cherts of the Barberton greenstone belt interpreted as products of submarine exhalative activity. *Journal of Geology* 93: 111–129.
- Parker TJ, Gorsline DS, Saunders RS, Pieri DC, and Schneeberger DM (1993) Coastal geomorphology of the Martian northern plains. *Journal of Geophysical Research* 98: 11061–11078.
- Parkes RJ, Cragg BA, Bale SJ, et al. (1994) Deep bacterial biosphere in Pacific Ocean sediments. *Nature* 371: 410–411.
- Parnell J, Lee P, Cockell CS, and Osinski GR (2004) Microbial colonization in impact generated hydrothermal sulphate deposits, Haughton impact structure, and implications for sulphates on Mars. *International Journal of Astrobiology* 3: 247–256.
- Pavlov AA, Kasting JF, Brown LL, Rages KA, and Freedman R (2001) Greenhouse warming by CH₄ in the atmosphere of early earth. *Journal of Geophysical Research* 105: 11981–11990.
- Pedersen K and Ekendahl S (1990) Distribution and activity of bacteria in deep granitic groundwaters of southeastern Sweden. *Microbial Ecology* 20: 37–52.
- Pirie NW (1973) On being the right size. *Annual Review of Microbiology* 27: 119–131.
- Prieto-Ballesteros O, Kargel JS, Fairén AG, Fernandez-Remolar DC, Dohm JM, and Amils R (2006) Interglacial clathrate destabilization on Mars: Possible contributing source of its atmospheric methane. *Geology* 34: 149–152.
- Purcell E (1977) Life at low Reynold's number. *American Journal of Physics* 45: 3–11.
- Rancourt DG, Thibault P-J, Mavrocordatos D, and Lamarche G (2005) Hydrous ferric oxide precipitation in the presence of nonmetabolizing bacteria: Constraints on the mechanism of a biotic effect. *Geochimica Cosmochimica Acta* 69: 553–577.
- Rasmussen B (2000) Filamentous microfossils in a 3,235-million-year-old volcanogenic massive sulphide deposit. *Nature* 405: 676–679.
- Reysenbach A-L and Shock E (2002) Merging genomes with geochemistry in hydrothermal ecosystems. *Science* 296: 1077–1082.
- Riding R (2000) Microbial carbonates: The geological record of calcified bacterial-algal mats and biofilms. *Sedimentology* 47: 179–214.
- Rivkina EM, Friedmann EI, McKay CP, and Gilichinsky DA (2000) Metabolic activity of permafrost bacteria below the freezing point. *Applied and Environmental Microbiology* 66: 3230–3233.
- Roden E, Sobolev D, Glazer B, and Luther G (2004) Potential for microscale bacterial Fe redox cycling at the aerobic-anaerobic interface. *Geomicrobiology Journal* 21: 379–391.
- Rogers JR, Bennett PC, and Choi WJ (1998) Feldspars as a source of nutrients for microorganisms. *American Mineralogist* 83: 1532–1540.
- Ryder G (2002) Mass influx in the ancient Earth–Moon system and benign implications for the origin of life on Earth. *Journal of Geophysical Research* 107: (doi 10.1029/2001JE001583).
- Rye R, Kuo PH, and Holland HD (1995) Atmospheric carbon dioxide concentrations before 2.2 billion years ago. *Nature* 378: 603–605.
- Schiaparelli GV (1878) La carta di Marte, Rome.
- Schidlowski M (1988) A 3800 million-year isotopic record of life from carbon in sedimentary rocks. *Nature* 333: 313–318.
- Schultze-Lam S, Ferris FG, Konhauser KO, and Wiese RG (1995) *In situ* silicification of an Icelandic hot spring microbial mat: Implications for microfossil formation. *Canadian Journal of Earth Sciences* 32: 2021–2026.
- Schwartz AW and Orgel LE (1985) Template-directed synthesis of novel, nucleic acid-like structures. *Science* 228: 585–587.
- Scow KM and Hutson J (1992) Effect of diffusion and sorption on the kinetics of biodegradation: Theoretical considerations. *Soil Science Society of America Journal* 56: 119–127.
- Singer PC and Stumm W (1970) Acidic mine drainage: The rate determining step. *Science* 176: 1121–1123.
- Sleep NH, Zahnle KJ, Kasting JF, and Morowitz HJ (1989) Annihilation of ecosystems by large asteroid impacts on the early Earth. *Nature* 342: 139–142.
- Southam G and Donald R (1999) A structural comparison of bacterial microfossils versus 'nanobacteria' and nanofossils. *Earth Science Review* 48: 251–264.
- Southam G, Brock C, Donald R, and Röstad A (2001) Pyrite discs in coal: Evidence for fossilized bacterial colonies. *Geology* 29: 47–50.
- Stetter KO, Fiala G, Huber G, Huber R, and Segerer A (1990) Hyperthermophilic microorganisms. *FEMS Microbiology Reviews* 75: 117–124.
- Stahl LJ (1994) Microbial mats: Ecophysiological interactions related to biogenic sediment stabilization. In: Krumböhl WE, Paterson DM, and Stahl LJ (eds.) *Biostabilization of Sediments*, pp. 41–54. Oldenburg: Biblioteks und Informations system der Carl von Ossietzky Universität.
- Stevens TO, McKinley JP, and Fredrickson JK (1993) Bacteria associated with deep, alkaline, anaerobic groundwaters in southeast Washington. *Microbial Ecology* 25: 35–50.
- Taunton AE, Welch SA, and Banfield JF (2000) Microbial controls on phosphate and lanthanide distributions during granite weathering. *Chemical Geology* 169: 371–382.
- Tazaki K, Ferris FG, Wiese RG, and Fyfe WS (1992) Iron and graphite associated with fossil bacteria in chert. *Chemical Geology* 95: 313–325.
- Thompson JB and Ferris FG (1990) Cyanobacterial precipitation of gypsum, calcite, and magnesite from natural alkaline lake water. *Geology* 18: 995–998.
- Thompson JB, Schultze-Lam S, Beveridge TJ, and Des Marais DJ (1997) Whiting events: Biogenic origin due to the photosynthetic activity of cyanobacterial picoplankton. *Limnology and Oceanography* 42: 133–141.
- Thomsen TR, Finster K, and Ramsing NB (2001) Biogeochemical and molecular signatures of anaerobic methane oxidation in a marine sediment. *Applied and Environmental Microbiology* 67: 1646–1656.
- Tice M and Lowe DR (2004) Photosynthetic microbial mats in the 3,416-Myr-old ocean. *Nature* 431: 549–552.
- Tuttle JH, Dugan CB, and Randles CI (1969) Microbial sulfate reduction and its potential utility as an acid mine water

- pollution abatement procedure. *Applied Microbiology* 17: 297–302.
- Tung HC, Bramall NE, and Price PB (2005) Microbial origin of excess methane in glacial ice and implications for life on Mars. *PNAS* 102: 18292–18296.
- Urrutia MM and Beveridge TJ (1994) Formation of fine-grained silicate minerals and metal precipitates by a bacterial surface (*Bacillus subtilis*) and the implications in the global cycling of silicon. *Chemical Geology* 116: 261–280.
- van Gemerden H (1986) Production of elemental sulphur by green and purple sulphur bacteria. *Archives of Microbiology* 146: 52–56.
- Wächtershäuser G (1988) Pyrite formation, the first energy source for life: A hypothesis. *Systematic and Applied Microbiology* 10: 207–210.
- Walsh MM and Lowe D (1999) Modes of accumulation of carbonaceous matter in the early Archaean: A petrographic and geochemical study of carbonaceous cherts from the Swaziland Supergroup. In: Lowe DR and Byerly GR (eds.) *Geologic Evolution of the Barberton Greenstone Belt, South Africa*. *Geol. Soc. Am Spec. Paper*, 329: 115–132.
- Wanger G, Onstott TC, and Southam G (2006) Structural and chemical characterization of a natural fracture surface from 2.8 kilometers below land surface: Biofilms in the deep subsurface. *Geomicrobiology Journal* 23: 443–452.
- Westall F, Boni L, and Guerzoni ME (1995) The experimental silicification of microbes. *Palaeontology* 38: 495–528.
- Westall F and Southam G (2006) Early life on Earth. In: Benn K, Mareschal J-C, and Condé K (eds.) *AGU Geophysical Monograph* 164, pp. 283–304. Washington, DC: American Geophysical Union.
- Westall F, de Vries ST, Nijman W, et al. (2006a) The 3.466 Ga Kitty's Gap Chert, an Early Archaean microbial ecosystem. In: Reimold WU and Gibson R (eds.) *Geological Society of America Special Paper 405: Processes on the Early Earth* pp. 105–131. Boulder, Co: Geological Society of America.
- Westall F, de Ronde CEJ, Southam G, et al. (2006b) Implications of a 3.472–3.333 Ga-old subaerial microbial mat from the Barberton greenstone belt, South Africa for the UV environmental conditions on the early Earth. *Philosophical Transactions of the Royal Society of London Series B* 361: 1857–1875.
- Woese CR, Kandler O, and Wheelis ML (1990) Towards a natural system of organisms: Proposal for the domains Archaea, Bacteria, and Eucarya. *PNAS* 87: 4576–4579.
- Youssef MI (1965) Genesis of bedded phosphates. *Economic Geology* 60: 590–600.
- Zaug AJ and Cech TR (1986) The intervening sequence RNA of *Tetrahymena* is an enzyme. *Science* 231: 470–475.

10.13 Giant Planets

T. Guillot, Observatoire de la Côte d'Azur, CNRS, Nice, France

D. Gautier, Observatoire de Paris, CNRS, Meudon, France

© 2007 Elsevier B.V. All rights reserved.

10.13.1	Introduction	439
10.13.2	Observations and Global Properties	440
10.13.2.1	Visual Appearances	440
10.13.2.2	Gravity Fields	440
10.13.2.3	Magnetic Fields	442
10.13.2.4	Atmospheric Compositions	442
10.13.2.4.1	Hydrogen and helium	442
10.13.2.4.2	Heavy elements	443
10.13.2.5	Energy Balance and Atmospheric Temperature Profiles	444
10.13.2.6	Atmospheric Dynamics: Winds and Weather	445
10.13.2.7	Moons and Rings	446
10.13.2.8	Extrasolar Planets	446
10.13.3	The Calculation of Interior and Evolution Models	447
10.13.3.1	Basic Equations	447
10.13.3.2	High-Pressure Physics and Equations of State	447
10.13.3.3	Heat Transport	449
10.13.3.4	The Contraction and Cooling Histories of Giant Planets	450
10.13.3.5	Mass–Radius Relation	451
10.13.3.6	Rotation and the Figures of Planets	452
10.13.4	Interior Structures and Evolutions	453
10.13.4.1	Jupiter and Saturn	453
10.13.4.2	Uranus and Neptune	455
10.13.4.3	Irradiated Giant Planets	456
10.13.5	Implications for Planetary Formation Models	459
10.13.6	Future Prospects	460
	References	461

10.13.1 Introduction

In our solar system, four planets stand out for their sheer mass and size. Jupiter, Saturn, Uranus, and Neptune indeed qualify as ‘giant planets’ because they are larger than any terrestrial planet and much more massive than all other objects in the solar system, except the Sun, put together (Figure 1). Because of their gravitational might, they have played a key role in the formation of the solar system, tossing around many objects in the system, preventing the formation of a planet in what is now the asteroid belt, and directly leading to the formation of the Kuiper Belt and Oort Cloud. They also retain some of the gas (in particular hydrogen and helium) that was present when the Sun and its planets formed

and are thus key witnesses in the search for our origins.

Because of a massive envelope mostly made of hydrogen helium, these planets are ‘fluid’, with no solid or liquid surface. In terms of structure and composition, they lie in between stars (gaseous and mostly made of hydrogen and helium) and smaller terrestrial planets (solid and liquid and mostly made of heavy elements), with Jupiter and Saturn being closer to the former and Uranus and Neptune to the latter.

The discovery of many extrasolar planets of masses from a few hundreds down to a few Earth masses and the possibility to characterize them by the measurement of their mass and size prompts a more general definition of giant planets. For this review,

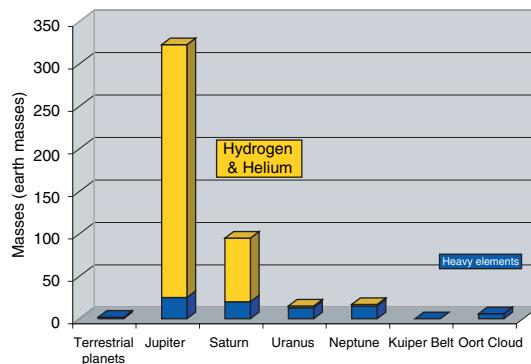


Figure 1 An inventory of hydrogen and helium and all other elements ('heavy elements') in the solar system excluding the Sun (the Sun has a total mass of $332\,960\,M_{\oplus}$, including about $5000\,M_{\oplus}$ in heavy elements, $1\,M_{\oplus}$ being the mass of the Earth). The precise amount of heavy elements in Jupiter ($10\text{--}40\,M_{\oplus}$) and Saturn ($20\text{--}30\,M_{\oplus}$) is uncertain (see Section 10.13.4.1).

we will adopt the following: "a giant planet is a planet mostly made of hydrogen and helium and too light to ignite deuterium fusion." This is purposely relatively vague – depending on whether the inventory is performed by mass or by atom or molecule, Uranus and Neptune may be included or left out of the category. Note that Uranus and Neptune are indeed relatively different in structure than Jupiter and Saturn and are generally referred to as 'ice giants', due to an interior structure that is consistent with the presence of mostly 'ices' (a mixture formed from the condensation in the protoplanetary disk of low-refractivity materials such as H_2O , CH_4 , and NH_3 , and brought to the high-pressure conditions of planetary interiors – see below).

Globally, this definition encompasses a class of objects that have similar properties (in particular, a low viscosity and a non-negligible compressibility) and inherited part of their material directly from the same reservoir as their parent star. These objects can thus be largely studied with the same tools, and their formation is linked to that of their parent star and the fate of the circumstellar gaseous disk present around the young star.

We will hereafter present some of the key data concerning giant planets in the solar system and outside. We will then present the theoretical basis for the study of their structure and evolution. On this basis, the constraints on their composition will be discussed and analyzed in terms of consequences for the models of planet formation.

10.13.2 Observations and Global Properties

10.13.2.1 Visual Appearances

In spite of its smallness, the sample of four giant planets in our solar system exhibits a large variety of appearances, shapes, colors, variability, etc. As shown in **Figure 2**, all four giant planets are flattened by rotation and exhibit a more or less clear zonal wind pattern, but the color of their visible atmosphere is very different (this is due mostly to minor species in the high planetary atmosphere), their clouds have different compositions (ammonia for Jupiter and Saturn, methane for Uranus and Neptune) and depths, and their global meteorology (number of vortices, long-lived anticyclones such as Jupiter's Great Red Spot, presence of planetary-scale storms, convective activity) is different from one planet to the next.

We can presently only wonder about what is in store for us with extrasolar giant planets since we cannot image them. But with orbital distances that can be as close as 0.02 AU, a variety of masses, sizes, and parent stars, we should expect to be surprised!

10.13.2.2 Gravity Fields

The mass of our giant planets can be obtained with great accuracy from the observation of the motions of their natural satellites: 317.834, 95.161, 14.538, and 17.148 times the mass of the Earth ($1M_{\oplus} = 5.97369 \times 10^{27}\,\text{g}$) for Jupiter, Saturn, Uranus and Neptune, respectively. More precise measurements of their gravity field can be obtained through the analysis of the trajectories of space crafts during flyby, especially when they come close to the planet and preferably in a near-polar orbit. The gravitational field thus measured departs from a purely spherical function due to the planets' rapid rotation. The measurements are generally expressed by expanding the



Figure 2 Photomontage from images of Voyager 2 (Jupiter, Uranus, and Neptune) and Cassini (Saturn). The planets are shown to scale, with their respective axial inclinations.

Table 1 Characteristics of the gravity fields and radii

	Jupiter	Saturn	Uranus	Neptune
$M \times 10^{-26}$ (kg)	18.986112(15) ^a	5.684640(30) ^b	0.8683205(34) ^c	1.0243542(31) ^d
$R_{\text{eq}} \times 10^{-7}$ (m)	7.1492(4) ^e	6.0268(4) ^f	2.5559(4) ^g	2.4766(15) ^g
$R_{\text{pol}} \times 10^{-7}$ (m)	6.6854(10) ^e	5.4364(10) ^f	2.4973(20) ^g	2.4342(30) ^g
$\bar{R} \times 10^{-7}$ (m)	6.9894(6) ^h	5.8210(6) ^h	2.5364(10) ⁱ	2.4625(20) ^j
$\bar{\rho} \times 10^{-3}$ (kg m ⁻³)	1.3275(4)	0.6880(2)	1.2704(15)	1.6377(40)
$J_2 \times 10^2$	1.4697(1) ^a	1.632425(27) ^b	0.35160(32) ^c	0.3539(10) ^d
$J_4 \times 10^4$	-5.84(5) ^a	-9.397(28) ^b	-0.354(41) ^c	-0.28(22) ^d
$J_6 \times 10^4$	0.31(20) ^a	0.867(97) ^b		
$P_{\omega} \times 10^{-4}$ (s)	3.57297(41) ^j	3.83624(47)? ^{j, k}	6.206(4) ^l	5.800(20) ^m
q	0.08923(5)	0.15491(10)	0.02951(5)	0.02609(23)
C/MR_{eq}^2	0.258	0.220	0.230	0.241

^aCampbell and Synnott (1985).^bJacobson *et al.* (2006).^cAnderson *et al.* (1987).^dTyler *et al.* (1989).^eLindal *et al.* (1981).^fLindal *et al.* (1985).^gLindal (1992).^hFrom fourth order figure theory.ⁱ($2R_{\text{eq}} + R_{\text{pol}}$)/3 (Clairaut's approximation).^jDavies *et al.* (1986).^kThis measurement from the Voyager era is now in question and values up to 38 826 s have been proposed (see Section 10.13.2.3).^lWarwick *et al.* (1986).^mWarwick *et al.* (1989).

The numbers in parentheses are the uncertainty in the last digits of the given value. The value of the gravitational constant used to calculate the masses of Jupiter and Saturn is $G = 6.67259 \times 10^{-11}$ N m² kg⁻² (Cohen and Taylor, 1987). The values of the radii, density, and gravitational moments correspond to the one bar pressure level (1 bar = 10⁵ Pa).

components of the gravity field on Legendre polynomials P_i of progressively higher orders:

$$V_{\text{ext}}(r, \theta) = -\frac{GM}{r} \left\{ 1 - \sum_{i=1}^{\infty} \left(\frac{R_{\text{eq}}}{r} \right)^i J_i P_i(\cos \theta) \right\} \quad [1]$$

where $V_{\text{ext}}(r, \theta)$ is the gravity field evaluated outside the planet at a distance r and colatitude θ , R_{eq} is the equatorial radius, and J_i are the gravitational moments. Because the giant planets are very close to hydrostatic equilibrium the coefficients of even order are the only ones that are not negligible. We will see how these gravitational moments, as listed in **Table 1**, help us constrain the planets' interior density profiles.

Table 1 also indicates the radii obtained with the greatest accuracy by radio-occultation experiments. An important consequence obtained is the fact that these planets have low densities, from 0.688 g cm⁻³ for Saturn to 1.64 g cm⁻³ for Neptune, to be compared with densities of 3.9–5.5 g cm⁻³ for the terrestrial planets. Considering the compression that strongly increases with mass, one is led naturally to the conclusion that these planets contain an important proportion of light materials including hydrogen and helium. It also implies that Uranus and Neptune

which are less massive must contain a relatively larger proportion of heavy elements than Jupiter and Saturn. This may lead to a subclassification between the hydrogen–helium giant planets Jupiter and Saturn, and the ‘ice giants’ or ‘sub giants’ Uranus and Neptune.

The planets are also relatively fast rotators, with periods of ~10 h for Jupiter and Saturn, and ~17 h for Uranus and Neptune. The fact that this fast rotation visibly affects the figure (shape) of these planets is seen by the significant difference between the polar and equatorial radii. It also leads to gravitational moments that differ significantly from a null value. However, it is important to stress that there is no unique rotation frame for these fluid planets: atmospheric zonal winds imply that different latitudes rotate at different velocities (see Section 10.13.2.6), and the magnetic field provides another rotation period. Because the latter is tied to the deeper levels of the planet, it is believed to be more relevant when interpreting the gravitational moments. The rotation periods listed in **Table 1** hence correspond to that of the magnetic field. The case of Saturn appears to be complex and is discussed in the next section.

10.13.2.3 Magnetic Fields

Like the Earth, the Sun, and Mercury, our four giant planets possess their own magnetic fields, as shown by the Voyager 2 measurements. The structures of these magnetic fields are very different from one planet to another (see Chapter 10.07) and the dynamo mechanism that generates them is believed to be related to convection in their interior but is otherwise essentially unknown (see Stevenson (1983) and Chapter 10.08 for a review).

The magnetic field \mathbf{B} is generally expressed in the form of a development in spherical harmonics of the scalar potential W , such that $\mathbf{B} = -\nabla W$:

$$W = a \sum_{n=1}^{\infty} \left(\frac{a}{r}\right)^{n+1} \sum_{m=0}^n \left\{ g_n^m \cos(m\phi) + b_n^m \sin(m\phi) \right\} \\ \times P_n^m(\cos \theta) \quad [2]$$

where r is the distance to the planet's center, a its radius, θ the colatitude, ϕ the longitude, and P_n^m the associated Legendre polynomials. The coefficients g_n^m and b_n^m are the magnetic moments that characterize the field. They are expressed in magnetic field units.

One can show that the first coefficients of relation [2] (for $n=0$ and $n=1$) correspond to the potential of a magnetic dipole such that $W = \mathbf{M} \cdot \mathbf{r}/r^3$ of moment:

$$M = a^3 \left\{ (g_1^0)^2 + (g_1^1)^2 + (b_1^1)^2 \right\}^{1/2} \quad [3]$$

Jupiter and Saturn have magnetic fields of essentially dipolar nature, of axis close to the rotation axis (g_1^0 is much larger than the other harmonics); Uranus and Neptune have magnetic fields that are intrinsically much more complex. To provide an idea of the intensity of the magnetic fields, the value of the dipolar moments for the four planets are 4.27 Gauss R_J^3 , 0.21 Gauss R_S^3 , 0.23 Gauss R_U^3 , 0.133 Gauss R_N^3 , respectively (Connerney *et al.*, 1982; Acuna *et al.*, 1983; Ness *et al.*, 1986, 1989).

A true surprise from Voyager that has been confirmed by the Cassini-Huygens mission is that Saturn's magnetic field is axisymmetric 'to the limit of the measurement accuracy': Saturn's magnetic and rotation axes are perfectly aligned. Voyager measurements indicated nevertheless a clear signature in the radio signal at 10 h 39 min 22 s believed to be a consequence of the rotation of the magnetic field. Determinations of a magnetic anomaly and new measurements by Cassini have since considerably blurred the picture, and the interpretation of the measurements have become unclear, with a deep rotation period evaluated between that of Voyager and as slow as 10 h 47 min 6 s (Gurnett *et al.*, 2005;

Giampieri *et al.*, 2006) (see also Galopeau and Lecacheux, 2000; Cecconi and Zarka, 2005). Note that models discussed hereafter have not yet included this additional uncertainty.

10.13.2.4 Atmospheric Compositions

In fluid planets, the distinction between the atmosphere and the interior is not obvious. We name 'atmosphere' the part of the planet which can directly exchange radiation with the exterior environment. This is also the part which is accessible by remote sensing. It is important to note that the continuity between the atmosphere and the interior does not guarantee that compositions measured in the atmosphere can be extrapolated to the deep interior, even in a fully convective environment: processes such as phase separations (e.g. Salpeter, 1973; Stevenson and Salpeter, 1977b; Fortney and Hubbard, 2003), phase transitions (e.g., Hubbard, 1989), chemical reactions (e.g., Fegley and Lodders, 1994) can occur and decouple the surface and interior compositions. Furthermore, imperfect mixing may also occur, depending on the initial conditions (e.g., Stevenson, 1985).

The conventional wisdom is however that these processes are limited to certain species (e.g., helium) or that they have a relatively small impact on the global abundances, so that the hydrogen–helium envelopes may be considered relatively uniform, from the perspective of the global abundance in heavy elements. We first discuss measurements made in the atmosphere before inferring interior compositions from interior and evolution models.

10.13.2.4.1 Hydrogen and helium

The most important components of the atmospheres of our giant planets are also among the most difficult to detect: H₂ and He have a zero dipolar moment and hence absorb very inefficiently visible and infrared light. Their infrared absorption becomes important only at high pressures when collision-induced absorption becomes significant (e.g., Borysow *et al.*, 1997). On the other hand, lines due to electronic transitions correspond to very high altitudes in the atmosphere, and bear little information on the structure of the deeper levels. The only robust result concerning the abundance of helium in a giant planet is by *in situ* measurement by the Galileo probe in the atmosphere of Jupiter (von Zahn *et al.*, 1998). The helium mole fraction (i.e., number of helium atoms over the total number of species in a given volume) is $q_{\text{He}} = 0.1359 \pm 0.0027$. The helium mass-mixing ratio Y (i.e., mass of helium atoms over

total mass) is constrained by its ratio over hydrogen, $X/Y/(X+Y) = 0.238 \pm 0.05$. This ratio is by coincidence that found in the Sun's atmosphere, but because of helium sedimentation in the Sun's radiative zone, it was larger in the protosolar nebula: $Y_{\text{proto}} = 0.275 \pm 0.01$ and $(X+Y)_{\text{proto}} \approx 0.98$ (e.g., Bahcall *et al.*, 1995). Less helium is therefore found in the atmosphere of Jupiter than inferred to be present when the planet formed. We will discuss the consequences of this measurement later: let us mention that the explanation invokes helium settling due to a phase separation in the interiors of massive and cold giant planets.

Helium is also found to be depleted compared to the protosolar value in Saturn's atmosphere. However, in this case the analysis is complicated by the fact that Voyager radio-occultations apparently led to a wrong value. The current adopted value is now $Y = 0.18 - 0.25$ (Conrath and Gautier, 2000), in agreement with values predicted by interior and evolution models (Guillot, 1999b; Hubbard *et al.*, 1999). Finally, Uranus and Neptune are found to have near-protosolar helium-mixing ratios, but with considerable uncertainty.

10.13.2.4.2 Heavy elements

The abundance of other elements than hydrogen and helium (that we will call hereafter 'heavy elements') bears crucial information for the understanding of the processes that led to the formation of these planets.

The most abundant heavy elements in the envelopes of our four giant planets are O, C, N, S. It is possible to model the chemistry of gases in the tropospheres from the top of the convective zone down to the 2000 K temperature level (Fegley and Lodders, 1994). Models conclude that, whatever the initial composition in these elements of planetesimals which collapsed with hydrogen onto Jupiter and Saturn cores during the last phase of the planetary formation, C in the upper tropospheres of giant planets is mainly in the form of gaseous CH₄, N in the form of NH₃, S in the form of H₂S, and O in the form of H₂O. All these gases, but methane in Jupiter and Saturn, condense in the upper troposphere, but vaporize at deeper levels when the temperature increases. Interestingly enough, noble gases are not expected to condense even at the cold tropopause temperatures of Uranus and Neptune.

The mass spectrometer aboard the Galileo atmospheric probe has performed *in situ* measurements of Ar, Kr, Xe, CH₄, NH₃, H₂S, and H₂O in the troposphere of Jupiter. C, N, and S were found to be oversolar by a factor 3–4 (Wong *et al.*, 2004), which

was not unexpected because condensation of nebula gases results in enriching icy grains and planetesimals. The surprise came from Ar, Kr, Xe, which were expected to be solar because they are difficult to condense, but turned out to be oversolar by factors 2–4 (Owen *et al.*, 1999; Wong *et al.*, 2004). One exception among these enriched species was neon, which was found to be significantly undersolar, but was predicted to be so because of a capture by the falling helium droplets (Roulston and Stevenson, 1995). Another exception was water, but this molecule is affected by meteorological processes, and the probe was shown to have fallen into a dry region of Jupiter's atmosphere (Table 2).

Specifically, CH₄/H₂ has been found oversolar in the four giant planets: the C/H ratio corresponding to the measured abundances is always higher than the solar C/H ratio, and in fact appears to be increasing with distance to the Sun. C/H is 3, 7.5, 45, and 45 times solar, in Jupiter, Saturn, Uranus, and Neptune, respectively. Note that the quoted enrichments are subject to changes when the solar abundances tables are revised, which happens surprisingly frequently.

Except for Jupiter, the determination of the NH₃ abundance is more uncertain than that of CH₄ because it is model dependent. It is derived from fitting microwave spectra of giant planets which exhibit a continuum opacity, more difficult to model than absorption spectral lines. However, the N/H enrichment seems to be, so far, fairly constant from a planet to another, around a factor 2, so that C/N is higher in Saturn than in Jupiter, and still higher in Uranus and Neptune.

H₂S has been measured *in situ* in Jupiter, but in the three other giant planets its large abundance is derived from the requirement to deplete NH₃ at deeper levels than the saturation one. This scenario has been proposed a long time ago by Gulkis *et al.* (1978). It implies that S/H is substantially oversolar in Uranus and in Neptune.

H₂O is difficult to measure in all four giant planets because of its relatively deep condensation. It was hoped that the Galileo probe would provide a measurement of its deep abundance, but the probe fell into one of Jupiter's 5-microns hotspot, what is now believed to be a dry region mostly governed by downwelling motions (e.g., Showman and Ingersoll, 1998). As a result, and although the probe provided measurements down to 22 bars, well below water's canonical 5 bar cloud base, it is believed that this measurement of a water abundance equal to a fraction of the solar value is only a lower limit.

Table 2 Main gaseous components of heavy elements measured in the troposphere of giant planets

	Species	Mixing ratio/H ₂	References	Comments
Jupiter	CH ₄	(2.37 ± 0.37) × 10 ⁻³	Wong <i>et al.</i> (2004)	GPMS on Galileo ^a
	NH ₃	(6.64 ± 2.34) × 10 ⁻³	Wong <i>et al.</i> (2004)	<i>idem</i>
	H ₂ S	(8.9 ± 2.1) × 10 ⁻³	Wong <i>et al.</i> (2004)	<i>idem</i>
	H ₂ O	(4.9 ± 1.6) × 10 ⁻⁴	Wong <i>et al.</i> (2004)	<i>idem</i> ; region not well mixed ^b
	³⁶ Ar	(6.1 ± 1.2) × 10 ⁻⁶	Atreya <i>et al.</i> (1999)	<i>idem</i>
	⁸⁴ Kr	(1.84 ± 0.37) × 10 ⁻⁹	Atreya <i>et al.</i> (1999)	<i>idem</i>
	¹³² Xe	(4.9 ± 1.0) × 10 ⁻¹¹	Atreya <i>et al.</i> (1999)	<i>idem</i>
Saturn	CH ₄	(4.3 ± 1) × 10 ⁻³	Flasar <i>et al.</i> (2005)	CIRS on Cassini ^c
	NH ₃	(1 ± 1) × 10 ⁻⁴	Briggs and Sackett (1989)	Ground-based microwave ^d
	H ₂ S	(2.2 ± 0.3) × 10 ⁻⁴	Briggs and Sackett (1989)	<i>idem</i>
Uranus	CH ₄	(3.3 ± 1.1) × 10 ⁻²	Gautier <i>et al.</i> (1995)	Compilation from ground-based observations
	H ₂ S	(1 ± 1) × 10 ⁻⁴	Briggs and Sackett (1989)	Ground-based microwave ^d
Neptune	CH ₄	(3.3 ± 1.1) × 10 ⁻²	Gautier <i>et al.</i> (1995)	Compilation from ground-based observations
	H ₂ S	(7.5 ± 3.25) × 10 ⁻⁴	de Pater <i>et al.</i> (1991)	Ground-based microwave ^e
	H ₂ O	7.7 × 10(−1)	Lodders and Fegley (1994)	Inferred from CO ^f

^aGalileo Probe Mass Spectrometer aboard the atmospheric probe in Jupiter.^bThe signal stopped at the 22 bar levels prior to have reached a constant value. It is currently believed that the region where the probe made measurements was atypically dry and that the bulk abundance of H₂O in Jupiter has not been measured.^cComposite Infra Red Spectrometer aboard the Cassini spacecraft.^dGround-based measurements of the microwave continuum. The result is somewhat uncertain due to the difficulty to precisely estimate opacities of absorbing species.^eGround-based microwave measurements. An oversolar H₂S abundance is required to interpret the depletion of NH₃ in the upper troposphere.^fInferred from the microwave detection of CO in the troposphere. Note that the validity of the approach is questionned by Bézard *et al.* (2002). A large amount of water seems to be present anyway in the deep atmosphere of Neptune. The case of Uranus is still uncertain because it is not known so far if CO is present in the troposphere of the planet.

10.13.2.5 Energy Balance and Atmospheric Temperature Profiles

Jupiter, Saturn, and Neptune are observed to emit significantly more energy than they receive from the Sun (see **Table 3**). The case of Uranus is less clear. Its intrinsic heat flux F_{int} is significantly smaller than that of the other giant planets. Detailed modeling of its atmosphere however indicates that $F_{\text{int}} \gtrsim 60 \text{ erg cm}^{-2} \text{ s}^{-1}$ (Marley and McKay, 1999). With this caveat, all four giant planets can be said to emit

more energy than they receive from the Sun. Hubbard (1968) showed in the case of Jupiter that this can be explained simply by the progressive contraction and cooling of the planets.

A crucial consequence of the presence of an intrinsic heat flux is that it requires high internal temperatures ($\sim 10\,000 \text{ K}$ or more), and that consequently the giant planets are ‘fluid’ (not solid) (Hubbard, 1968; see also Hubbard *et al.*, 1995). Another consequence is that they are essentially

Table 3 Energy balance as determined from Voyager IRIS data

	Jupiter	Saturn	Uranus	Neptune
Absorbed power (10^{16} Js^{-1})	50.14 ± 2.48	11.14 ± 0.50	0.526 ± 0.037	0.204 ± 0.019
Emitted power (10^{16} Js^{-1})	83.65 ± 0.84	19.77 ± 0.32	0.560 ± 0.011	0.534 ± 0.029
Intrinsic power (10^{16} Js^{-1})	33.5 ± 2.6	8.63 ± 0.60	$0.034^{+0.038}_{-0.034}$	0.330 ± 0.035
Intrinsic flux ($\text{Js}^{-1} \text{ m}^{-2}$)	5.44 ± 0.43	2.01 ± 0.14	$0.042^{+0.047}_{-0.042}$	0.433 ± 0.046
Bond albedo	0.343 ± 0.032	0.342 ± 0.030	0.300 ± 0.049	0.290 ± 0.067
Effective temperature (K)	124.4 ± 0.3	95.0 ± 0.4	59.1 ± 0.3	59.3 ± 0.8
1-bar temperature ^a (K)	165 ± 5	135 ± 5	76 ± 2	72 ± 2

^aLindal (1992).

Adapted Pearl JC and Conrath BJ (1991) The albedo, effective temperature, and energy balance of Neptune, as determined from Voyager data. *Journal of Geophysical Research* 96: 18921.

convective, and that their interior temperature profiles are close to ‘adiabats’. We will come back to this in more detail.

The deep atmospheres (more accurately tropospheres) of the four giant planets are indeed observed to be close to adiabats, a result first obtained by spectroscopic models (Trafton, 1967), then verified by radio-occultation experiments by the Voyager space-crafts, and by the *in situ* measurement from the Galileo probe (**Figure 3**). The temperature profiles show a temperature minimum, in a region near 0.2 bar called the tropopause. At higher altitudes, in the stratosphere, the temperature gradient is negative (increasing with decreasing pressure). In the regions that we will be mostly concerned with, in the troposphere and in the deeper interior, the temperature always increases with depth. It can be noticed that the slope of the temperature profile in **Figure 3** becomes almost constant when the atmosphere becomes convective, at pressures of a few tens of bars, in the four giant planets.

It should be noted that the 1 bar temperatures listed in **Table 3** and the profiles shown in **Figure 3** are retrieved from radio-occultation measurements using a helium-to-hydrogen ratio which, at least in the case of Jupiter and Saturn, was shown to be incorrect. The new values of Y are found to lead to increased temperatures by ~ 5 K in Jupiter and ~ 10 K in Saturn (see Guillot, 1999a). However, the Galileo probe found a 1 bar temperature of 166 K (Seiff *et al.*, 1998), and generally a good agreement with the Voyager radio-occultation profile with the wrong He/H₂ value.

When studied at low spatial resolution, it is found that all four giant planets, in spite of their

inhomogeneous appearances, have a rather uniform brightness temperature, with pole-to-equator latitudinal variations limited to a few kelvins (e.g., Ingersoll *et al.*, 1995). However, in the case of Jupiter, some small regions are known to be very different from the average of the planet. This is the case of hotspots, which cover about 1% of the surface of the planet at any given time, but contribute to most of the emitted flux at 5 μ m, due to their dryness (absence of water vapor) and their temperature brightness which can, at this wavelength, peak to 260 K.

10.13.2.6 Atmospheric Dynamics: Winds and Weather

The atmospheres of all giant planets are evidently complex and turbulent in nature. This can for example be seen from the mean zonal winds (inferred from cloud tracking), which are very rapidly varying functions of the latitude (see, e.g., Ingersoll *et al.*, 1995): while some of the regions rotate at the same speed as the interior magnetic field (in the so-called ‘system III’ reference frame), most of the atmospheres do not. Jupiter and Saturn both have superrotating equators (+100 and +400 m s⁻¹ in system III, for Jupiter and Saturn, respectively), Uranus and Neptune have sub-rotating equators, and superrotating high-latitude jets. Neptune, which receives the smallest amount of energy from the Sun has the largest peak-to-peak latitudinal variations in wind velocity: about 600 m s⁻¹. It can be noted that, contrary to the case of the strongly irradiated planets to be discussed later, the winds of Jupiter, Saturn, Uranus, and Neptune, are significantly slower than the planet itself under its own spin (from 12.2 for Jupiter to 2.6 km s⁻¹ for Neptune, at the equator).

The observed surface winds are believed to be related to motions in the planets’ interiors, which, according to the Taylor–Proudman theorem, should be confined by the rapid rotation to the plane perpendicular to the axis of rotation (e.g., Busse, 1978). Unfortunately, no convincing model is yet capable of modeling with sufficient accuracy both the interior and the surface layers.

Our giant planets also exhibit planetary-scale to small-scale storms with very different temporal variations. For example, Jupiter’s Great Red Spot is a 12 000 km-diameter anticyclone found to have lasted for at least 300 years (e.g., Simon-Miller *et al.*, 2002). Storms developing over the entire planet have even been observed on Saturn (Sanchez-Lavega *et al.*, 1996). Uranus and Neptune’s storm system has been shown to have been significantly altered since the Voyager

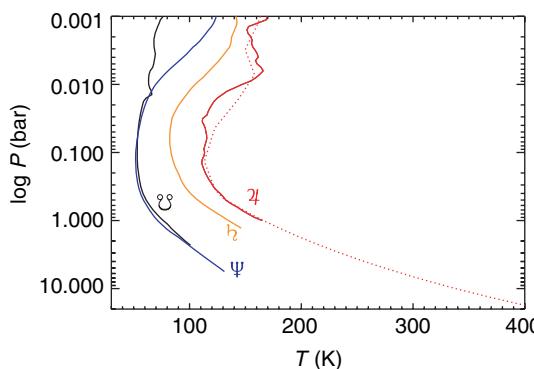


Figure 3 Atmospheric temperatures as a function of pressure for Jupiter, Saturn, Uranus, and Neptune, as obtained from Voyager radio-occultation experiments (see Lindal, 1992). The dotted line corresponds to the temperature profile retrieved by the Galileo probe, down to 22 bar and a temperature of 428 K (Seiff *et al.*, 1998).

era (Rages *et al.*, 2002; Hammel *et al.*, 2005). On Jupiter, small-scale storms related to cumulus-type cloud systems have been observed (e.g., Gierasch *et al.*, 2000; Hueso *et al.*, 2002), and lightning strikes have been monitored by Galileo (e.g., Little *et al.*, 1999). These represent only a small arbitrary subset of the work concerning the complex atmospheres of these planets.

It is tempting to extrapolate these observations to the objects outside our solar system as well. However, it is important to stress that an important component of the variability in the atmospheres of our giant planets is the presence of relatively abundant condensing chemical species: ammonia and water in the case of Jupiter and Saturn, and methane for Uranus and Neptune. These species can only condense in very cold atmospheres, thus providing latent heat to fuel important storms. Depending on their temperatures and compositions, extrasolar planets may or may not possess such important condensing species (e.g., Guillot, 1999b).

10.13.2.7 Moons and Rings

A discussion of our giant planets motivated by the opportunity to extrapolate the results to objects outside our solar system would be incomplete without mentioning the moons and rings that these planets all possess (see Chapters 10.09, 10.14, 10.15). First, the satellites/moons can be distinguished from their orbital characteristics as regular or irregular. The first ones have generally circular, prograde orbits. The latter tend to have eccentric, extended, and/or retrograde orbits.

These satellites are numerous: After the Voyager era, Jupiter was known to possess 16 satellites, Saturn to have 18, Uranus 20, and Neptune 8. Recent extensive observation programs have seen the number of satellites increase considerably, with a growing list of satellites presently reaching 62, 56, 27, and 13 for Jupiter, Saturn, Uranus, and Neptune, respectively. All of the new satellites discovered since Voyager are classified as irregular.

The presence of regular and irregular satellites is due in part to the history of planet formation. It is believed that the regular satellites have mostly been formed in the protoplanetary subnebulae that surrounded the giant planets (at least Jupiter and Saturn) at the time when they accreted their envelopes (cf. Chapter 10.14). On the other hand, the irregular satellites are thought to have been captured by the planet. This is for example believed to be the case of Neptune's largest moon, Triton, which has a retrograde orbit.

A few satellites stand out by having relatively large masses: it is the case of Jupiter's Io, Europa, Ganymede,

and Callisto; of Saturn's Titan; and of Neptune's Triton. Ganymede is the most massive of them, being about twice the mass of our Moon. However, compared to the mass of the central planet, these moons and satellites have very small weights: 10^{-4} and less for Jupiter, 1/4000 for Saturn, 1/25 000 for Uranus, and 1/4500 for Neptune. All these satellites orbit relatively closely to their planets. The farthest one, Callisto revolves around Jupiter in about 16 Earth days.

The four giant planets also have rings, whose material is probably constantly resupplied from their satellites. The ring of Saturn stands out as the only one directly visible with binoculars. In this particular case, its enormous area allows it to reflect a sizable fraction of the stellar flux arriving at Saturn, and makes this particular ring as bright as the planet itself. The occurrence of such rings would make the detection of extrasolar planets slightly easier, but it is yet unclear how frequent they can be, and how close to the stars rings can survive both the increased radiation and tidal forces.

10.13.2.8 Extrasolar Planets

Huge progresses have been made in the field of extrasolar planets since the detection of the first giant planet orbiting a solar-type star by Mayor and Queloz (1995). More than 200 planets are known at the time of this review, and importantly, 14 planets that transit their star at each orbital revolution have been identified (see Figure 4). These transiting planets are especially interesting because of the

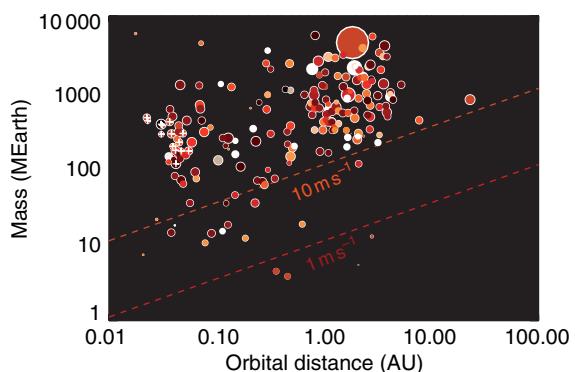


Figure 4 Masses and orbital distances of the extrasolar planets discovered by 2006. The size of the symbols is proportional to the mass of the parent star (from 0.1 to 4 stellar masses). The color (from white to red and black) is proportional to the stellar metallicity. Stars with metallicities $[Fe/H] < 0.2$ are shown in white. The radial velocimetry thresholds at 1 and 10 m s^{-1} , encompassing the detection limits of most current surveys, are indicated as dashed lines. 'Transiting' planets are highlighted with white crosses.

possibility to measure both their mass and size and thus obtain constraints on their global composition.

In spite of their particular location just a few stellar radii away from their stars, the transiting planets that have been discovered bear some resemblance with their solar system cousins in the sense that they are also mostly made of hydrogen and helium (e.g., Burrows *et al.*, 2000; Guillot, 2005; Baraffe *et al.*, 2005). They are, however, much hotter due to the intense irradiation that they receive.

Although obtaining direct informations on these planets represents a great observational challenge, several key steps have been accomplished: atomic sodium, predicted to be detectable (Seager and Sasselov, 2000), has indeed been detected by transit spectroscopy around one planet (Charbonneau *et al.*, 2002). Hydrodynamically escaping hydrogen, oxygen, and carbon have also been detected from the same planet (Vidal-Madjar *et al.*, 2003, 2004). The measurement of the ‘secondary’ eclipse of several planets by the Spitzer satellite allowed a constraint on spectral and hence thermal properties of the planetary atmospheres (Charbonneau *et al.*, 2005; Deming *et al.*, 2005). Recently, the light curve of a nontransiting planet was detected in the infrared, also with Spitzer, providing preliminary indications of a strong day/night temperature variation (Harrington *et al.*, 2006) perhaps even larger than predicted (see Showman and Guillot, 2002).

Obviously, there is a big potential for growth in this young field, and the comparison between fine observations made for giant planets in our solar system and the more crude, but also more statistically significant data obtained for planets around other stars promises to be extremely fruitful to better understand these objects.

10.13.3 The Calculation of Interior and Evolution Models

10.13.3.1 Basic Equations

The structure and evolution of a giant planet is governed by the following hydrostatic, thermodynamic, mass conservation, and energy conservation equations:

$$\frac{\partial P}{\partial r} = -\rho g \quad [4]$$

$$\frac{\partial T}{\partial r} = \frac{\partial P}{\partial r} \frac{T}{P} \nabla_T \quad [5]$$

$$\frac{\partial m}{\partial r} = 4\pi r^2 \rho \quad [6]$$

$$\frac{\partial L}{\partial r} = 4\pi r^2 \rho \left(\dot{\epsilon} - T \frac{\partial S}{\partial t} \right) \quad [7]$$

where P is the pressure, ρ the density, and $g = Gm/r^2$ the gravity (m is the mass, r the radius, and G the gravitational constant). The temperature gradient $\nabla_T \equiv (d \ln T / d \ln P)$ depends on the process by which the internal heat is transported. L is the intrinsic luminosity, t the time, S the specific entropy (per unit mass), and $\dot{\epsilon}$ accounts for the sources of energy due, for example, to radioactivity or more importantly nuclear reactions. Generally it is a good approximation to assume $\dot{\epsilon} \sim 0$ for objects less massive than $\sim 13 M_J$, that is, too cold to even burn deuterium (but we will see that in certain conditions this term may be useful, even for low-mass planets).

The boundary condition at the center is trivial: $r=0$; ($m=0$, $L=0$). The external boundary condition is more difficult to obtain because it depends on how energy is transported in the atmosphere. One possibility is to use the Eddington approximation, and to write (e.g., Chandrasekhar, 1939): $r=R$; ($T_0 = T_{\text{eff}}$, $P_0 = 2/3g/\kappa$), where T_{eff} is the effective temperature (defined by $L = 4\pi R \sigma T_{\text{eff}}^4$, with σ being the Stephan-Boltzmann constant), and κ is the opacity. Note for example that in the case of Jupiter $T_{\text{eff}} = 124$ K, $g = 26$ m s⁻², and $\kappa \approx 5 \times 10^{-3}$ (P/1 bar) m² kg⁻¹. This implies $P_0 \approx 0.2$ bar (20 000 Pa), which is actually close to Jupiter’s tropopause, where $T \approx 110$ K.

More generally, one has to use an atmospheric model relating the temperature and pressure at a given level to the radius R , intrinsic luminosity L , and incoming stellar luminosity L_{*p} : $r=R$; ($T_0 = T_0(R, L, L_{*p})$, $P_0 = P_0(R, L, L_{*p})$). P_0 is chosen to satisfy the condition that the corresponding optical depth at that level should be much larger than unity. If the stellar flux is absorbed mostly in a convective zone, then the problem can be simplified by using $T_0(R, L, L_{*p}) \approx T_0(R, L + L_{*p}, 0)$ (e.g., Hubbard, 1977). An example of such a model is described by Saumon *et al.* (1996) and Hubbard *et al.* (2002) and is used hereafter to model the planets in the low-irradiation limit.

10.13.3.2 High-Pressure Physics and Equations of State

In terms of pressures and temperatures, the interiors of giant planets lie in a region for which accurate equations of state (EOS) are extremely difficult to calculate. This is because both molecules, atoms, and ions can all coexist, in a fluid that is partially degenerate (free electrons have energies that are

determined both by quantum and thermal effects) and partially coupled (coulomb interactions between ions are not dominant but must be taken into account). The presence of many elements and their possible interactions further complicate matters. For lack of space, this section will mostly focus on hydrogen whose EOS has seen the most important developments in recent years. A phase diagram of hydrogen (**Figure 5**) illustrates some of the important phenomena that occur in giant planets.

The photospheres of giant planets are generally relatively cold (50–3000 K) and at low pressure (0.1–10 bar, or 10^4 – 10^6 Pa), so that hydrogen is in molecular form and the perfect gas conditions apply. As one goes deeper into the interior, hydrogen and helium progressively become fluid. (The perfect gas relation tends to underestimate the pressure by 10% or more when the density becomes larger than about 0.02 g cm^{-3} ($P \gtrsim 1 \text{ kbar}$ in the case of Jupiter)).

Characteristic interior pressures are considerably larger however: as implied by eqns [4] and [6], $P_c \approx GM^2/R^4$, of the order of 10–100 Mbar for Jupiter and Saturn. At these pressures and the

corresponding densities, the Fermi temperature T_F is larger than 10^5 K. This implies that electrons are degenerate. **Figure 5** shows that inside Jupiter, Saturn, the extrasolar planet HD 209458 b, but also for giant planets in general for most of their history, the degeneracy parameter $\theta = T/T_F$ is between 0.1 and 0.03. Therefore, the energy of electrons in the interior is expected to be only slightly larger than their nonrelativistic, fully degenerate limit: $u_e \geq 3/5kT_F = 15.6(\rho/\mu_e)^{2/3} \text{ eV}$, where k is Boltzmann's constant, μ_e is the number of electrons per nucleon, and ρ is the density in g cm^{-3} . For pure hydrogen, when the density reaches $\sim 0.8 \text{ g cm}^{-3}$, the average energy of electrons becomes larger than hydrogen's ionization potential, even at zero temperature: hydrogen pressure-ionizes and becomes metallic. This molecular-to-metallic transition occurs near Mbar pressures, but exactly how this happens remains unclear because of the complex interplay of thermal, coulomb, and degeneracy effects (in particular, whether hydrogen metallizes into an atomic state H^+ – as suggested in **Figure 5** – or first metallizes in the molecular state H_2 remains to be clarified).

Recent laboratory measurements on fluid deuterium have been able to reach pressures above $\gtrsim 1$ Mbar, and provide new data in a region where the EOS remains most uncertain. Gas-guns experiments have been able to measure the reshock temperature (Holmes *et al.*, 1995), near $T \sim 5000$ K, $P \sim 0.8$ Mbar, and a rise in the conductivity of molecular hydrogen up to $T \sim 3000$ K, $P \sim 1.4$ Mbar, a sign that metallicity may have been reached (Weir *et al.*, 1996). The following few years have seen the development of laser-induced shock compression (da Silva *et al.*, 1997; Collins *et al.*, 1998), pulsed-power shock compression (Knudson *et al.*, 2004), and convergent shock wave experiments (Belov *et al.*, 2002; Boriskov *et al.*, 2005) in a high-pressure ($P = 0.3$ – 4 Mbar) high-temperature ($T \sim 6000$ – 10^5 K) regime. Unfortunately, experimental results along the principal Hugoniot of deuterium do not agree in this pressure range. Laser-compression data give a maximum compression of ~ 6 while both the pulsed power-compression experiments and the convergent shock wave experiments find a value of ~ 4 . Models that are partly calibrated with experimental data (Saumon *et al.*, 1995; Ross, 1998; Ross and Yang, 2001) obtain a generally good agreement with the laser-compression data. However, the fact that independent models based on first principles (Militzer *et al.*, 2001; Desjarlais, 2003; Bonev *et al.*, 2004) yield low compressions strongly favors this solution.

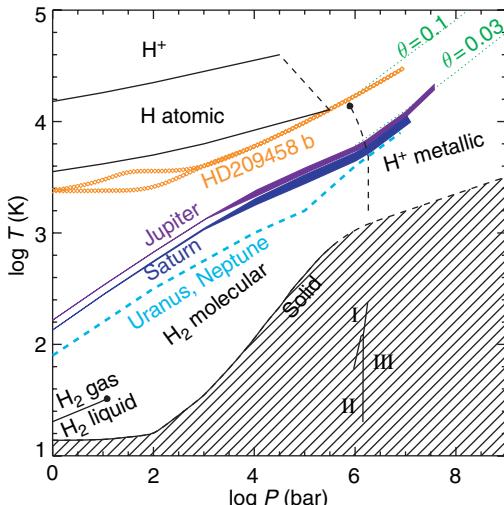


Figure 5 Phase diagram for hydrogen with the main-phase transitions occurring in the fluid or gas phase. The temperature-pressure profiles for Jupiter, Saturn, Uranus, Neptune, and the exoplanet HD 209458 b are shown. The dashed nearly vertical line near 1 Mbar is indicative of the molecular-to-metallic transition (here it represents the so-called plasma-phase transition as calculated by Saumon *et al.* (1995)). The region in which hydrogen is in solid phase (Datchi *et al.*, 2000; Gregoryanz *et al.*, 2003) is represented as a hatched area. The three phases (I, II, III) of solid hydrogen are shown (see Mao and Hemley, 1994). Values of the degeneracy parameter θ are indicated as dotted lines to the upper right corner of the figure.

The question of the existence of a first-order molecular-to-metallic transition of hydrogen (i.e., both molecular dissociation and ionization occur simultaneously and discontinuously at the so-called plasma phase transition, or PPT) remains however. The critical line shown in [Figure 5](#) corresponds to calculations by [Saumon et al. \(1995\)](#), but may be caused by artefacts in the free-energy calculation. Recent density functional theory (DFT) simulations by [Bonev et al. \(2004\)](#) indicate the possibility of a first-order liquid–liquid transition but other path-integral calculations ([Militzer et al., 2001](#)) do not. It is crucial to assess the existence of such a PPT because it would affect both convection and chemical composition in the giant planets.

A clear result from [Figure 5](#) at least is that, as first shown by [Hubbard \(1968\)](#), the interiors of the hydrogen–helium giant planets are ‘fluid’, whatever their age: of course, they avoid the critical point for the liquid gas transition in hydrogen and helium, at very low temperatures, but they also lie comfortably above the solidification lines for hydrogen and helium. (An ‘isolated’ Jupiter should begin partial solidification only after at least $\sim 10^3$ Ga of evolution.) They are considered to be fluid because at the high pressures and relatively modest temperatures in their interiors, coulomb interactions between ions play an important role in the EOS and yield a behavior that is more reminiscent of that of a liquid than that of a gas, contrary to what is the case in, for example, solar-like stars. For Uranus and Neptune, the situation is actually more complex because at large pressures they are not expected to contain hydrogen, but numerical simulations show that ices in their interior should be fluid as well ([Cavazzoni et al., 1999](#)).

Models of the interiors of giant planets require thermodynamically consistent EOSs calculated over the entire domain of pressure and temperature spanned by the planets during their evolution. Elements other than hydrogen, most importantly helium, should be consistently included. Such a calculation is a daunting task, and the only recent attempt at such an astrophysical EOS for substellar objects is that by [Saumon et al. \(1995\)](#). Another set of EOSs reproducing either the high- or low-compression results was calculated by [Saumon and Guillot \(2004\)](#) specifically for the calculation of present-day models of Jupiter and Saturn.

These EOSs have so far included other elements (including helium), only in a very approximative way, that is, with EOSs for helium and heavy elements that are based on interpolations between

somewhat ideal regimes, using an additive volume law, and neglecting the possibility of existence of phase separations (see [Hubbard et al., 2002; Guillot et al., 2004](#), for further discussions).

10.13.3.3 Heat Transport

Giant planets possess hot interiors, implying that a relatively large amount of energy has to be transported from the deep regions of the planets to their surface. This can either be done by radiation, conduction, or, if these processes are not sufficient, by convection. Convection is generally ensured by the rapid rise of the opacity with increasing pressure and temperature. At pressures of a bar or more and relatively low temperatures (less than 1000 K), the three dominant sources of opacities are water, methane, and collision-induced absorption by hydrogen molecules.

However, in the intermediate temperature range between ~ 1200 and 1500 K, the Rosseland opacity due to the hydrogen and helium absorption behaves differently: the absorption at any given wavelength increases with density, but because the temperature also rises, the photons are emitted at shorter wavelengths, where the monochromatic absorption is smaller. As a consequence, the opacity can decrease. This was shown by [Guillot et al. \(1994\)](#) to potentially lead to the presence of a deep radiative zone in the interiors of Jupiter, Saturn, and Uranus.

This problem must however be reanalyzed in the light of recent observations and analyses of brown dwarfs. Their spectra show unexpectedly wide sodium and potassium absorption lines (see [Burrows et al., 2000](#)), in spectral regions where hydrogen, helium, water, methane and ammonia are relatively transparent. It thus appears that the added contribution of these elements (if they are indeed present) would wipe out any radiative region at these levels ([Guillot et al., 2004](#)).

At temperatures above 1500 \sim 2000 K two important sources of opacity appear: (1) the rising number of electrons greatly enhances the absorption of H_2^- and H^- and (2) TiO, a very strong absorber at visible wavelengths is freed by the vaporization of $CaTiO_3$. Again, the opacity rises rapidly which ensures a convective transport of the heat. Still deeper, conduction by free electrons becomes more efficient, but the densities are found not to be high enough for this process to be significant, except perhaps near the central core (see [Hubbard, 1968; Stevenson and Salpeter, 1977b](#)).

While our giant planets seem to possess globally convective interiors, strongly irradiated extrasolar planets must develop a radiative zone just beneath the levels where most of the stellar irradiation is absorbed. Depending on the irradiation and characteristics of the planet, this zone may extend down to kbar levels, the deeper levels being convective. In this case, a careful determination of the opacities is necessary (but generally not possible) as these control the cooling and contraction of the deeper interior (see Ferguson *et al.*, 2005, for a discussion of opacities and tables for substellar atmospheres and interiors).

10.13.3.4 The Contraction and Cooling Histories of Giant Planets

The interiors of giant planets is expected to evolve with time from a high entropy, high θ value, hot initial state to a low entropy, low θ , cold degenerate state. The essential physics behind can be derived from the well-known virial theorem and the energy conservation which link the planet's internal energy E_i , gravitational energy E_g and luminosity through

$$\xi E_i + E_g = 0 \quad [8]$$

$$L = -\frac{\xi - 1}{\xi} \frac{dE_g}{dt} \quad [9]$$

where $\xi = \int_0^M 3(P/\rho)dm / \int_0^M u dm \approx <3P/\rho u>$, the brackets indicating averaging, and u is the specific internal energy. For a diatomic perfect gas, $\xi = 3.2$; for fully degenerate nonrelativistic electrons, $\xi = 2$.

Thus, for a giant planet or brown dwarf beginning its life mostly as a perfect H_2 gas, two-thirds of the energy gained by contraction is radiated away, one-third being used to increase E_i . The internal energy being proportional to the temperature, the effect is to heat up the planet. This represents the slightly counter-intuitive but well-known effect that a star or giant planet initially heats up while radiating a significant luminosity (e.g., Kippenhahn and Weigert, 1994).

Let us now move further in the evolution, when the contraction has proceeded to a point where the electrons have become degenerate. For simplicity, we will ignore coulomb interactions and exchange terms, and assume that the internal energy can be written as $E_i = E_{el} + E_{ion}$, and that furthermore $E_{el} \gg E_{ion}$ (θ is small). Because $\xi \approx 2$, we know that half of the gravitational potential energy is radiated away and half of it goes into internal energy. The problem is to decide how this energy is split into an electronic and

an ionic part. The gravitational energy changes with some average value of the interior density as $E_g \propto 1/R \propto \rho^{1/3}$. The energy of the degenerate electrons is essentially the Fermi energy: $E_{el} \propto \rho^{2/3}$. Therefore, $\dot{E}_{el} \approx 2(E_{el}/E_g)\dot{E}_g$. Using the virial theorem, this yields

$$\dot{E}_{el} \approx -\dot{E}_g \approx 2L \quad [10]$$

$$L \approx -\dot{E}_{ion} \propto -\dot{T} \quad [11]$$

The gravitational energy lost is entirely absorbed by the degenerate electrons, and the observed luminosity is due to the thermal cooling of the ions.

Several simplifications limit the applicability of this result (that would be valid in the white dwarf regime). In particular, the coulomb and exchange terms in the EOS introduce negative contributions that cannot be neglected. However, the approach is useful to grasp how the evolution proceeds: in its very early stages, the planet is very compressible. It follows a standard Kelvin–Helmoltz contraction. When degeneracy sets in, the compressibility becomes much smaller ($\alpha T \sim 0.1$, where α is the coefficient of thermal expansion), and the planet gets its luminosity mostly from the thermal cooling of the ions. The luminosity can be written in terms of a modified Kelvin–Helmoltz formula:

$$L \approx \eta \frac{GM^2}{R\tau} \quad [12]$$

where τ is the age, and η is a factor that hides most of the complex physics. In the approximation that coulomb and exchange terms can be neglected, $\eta \approx \theta/(\theta + 1)$. The poor compressibility of giant planets in their mature evolution stages implies that $\eta \ll 1$ ($\eta \sim 0.03$ for Jupiter): the luminosity is not obtained from the entire gravitational potential, but from the much more limited reservoir constituted by the thermal internal energy. Equation [12] shows that to first order, $\log L \propto -\log \tau$: very little time is spent at high luminosity values. In other words, the problem is (in most cases) weakly sensitive to initial conditions. However, it is to be noticed that with progresses in our capabilities to detect very young objects, that is, planets and brown dwarfs of only a few million years of age, the problem of the initial conditions does become important (Marley *et al.*, 2007).

Figure 6 shows more generally how giant planets, but also brown dwarfs and small stars see their luminosities evolve as a function of time. The $1/\tau$ slope is globally conserved, with some variations for brown

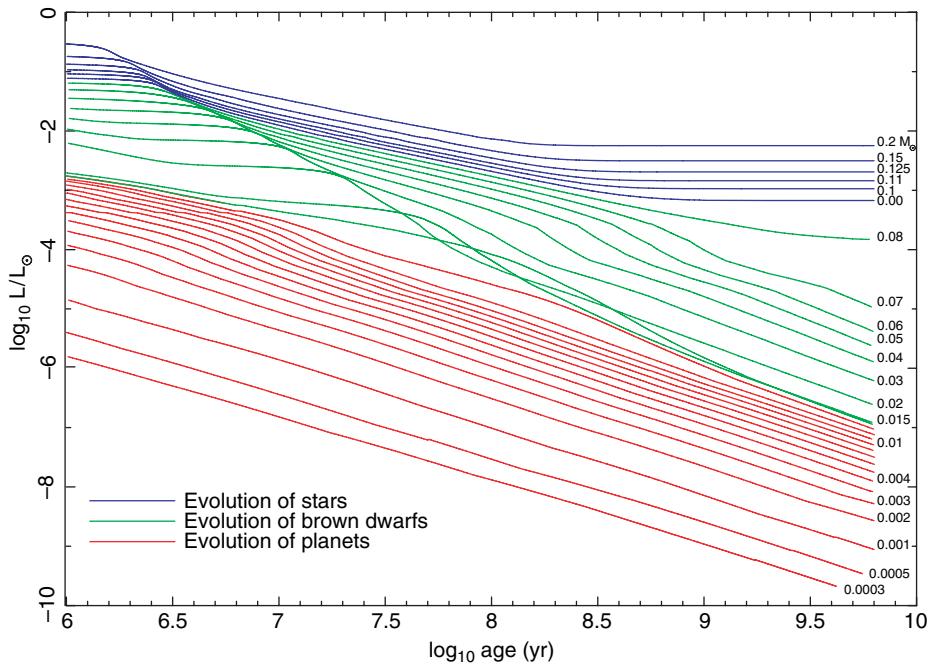


Figure 6 Evolution of the luminosity (in L_{\odot}) of solar-metallicity M dwarfs and substellar objects vs time (in yr) after formation. In this figure, ‘brown dwarfs’ are arbitrarily designated as those objects that burn deuterium, while those that do not are tentatively labelled ‘planets’. Stars are objects massive enough to halt their contraction due to hydrogen fusion. Each curve is labeled by its corresponding mass in M_{\odot} , with the lowest three corresponding to the mass of Saturn, half the mass of Jupiter, and the mass of Jupiter. From Burrows A, Marley M, Hubbard WB, et al. (1997) A nongray theory of extrasolar giant planets and brown dwarfs. *Astrophysical Journal* 491: 856–875.

dwarfs during the transient epoch of deuterium burning, and of course for stars, when they begin burning efficiently their hydrogen and settle on the main sequence: in that case, the tendency of the star to contract under the action of gravity is exactly balanced by thermonuclear hydrogen fusion.

10.13.3.5 Mass–Radius Relation

The relation between mass and radius has very fundamental astrophysical applications. Most importantly it allows one to infer the gross composition of an object from a measurement of its mass and radius. This is especially relevant in the context of the discovery of extrasolar planets with both radial velocimetry and the transit method, as the two techniques yield relatively accurate determination of M and R .

Figure 7 shows mass–radius relations for compact degenerate objects from giant planets to brown dwarfs and low-mass stars. The right-hand side of the diagram shows a rapid increase of the radius with mass in the stellar regime which is directly due to the

onset of stable thermonuclear reactions. In this regime, observations and theoretical models agree (see, however, Ribas (2006), for a more detailed discussion). The left-hand side of the diagram is obviously more complex, and this can be understood by the fact that planets have much larger variations in compositions than stars, and because external factors such as the amount of irradiation they receive do affect their contraction in a significant manner.

Let us first concentrate on isolated or nearly isolated gaseous planets. The black curves have a local maximum near $4 M_J$: at small masses, the compression is small so that the radius increases with mass. At large masses, degeneracy sets in and the radius decreases with mass.

This can be understood on the basis of polytropic models based on the assumption that $P = K\rho^{1+1/n}$, where K and n are constants. Because of degeneracy, a planet of large mass will tend to have $n \rightarrow 1.5$, while a planet a smaller mass will be less compressible ($n \rightarrow 0$). Indeed, it can be shown that in their inner 70 to 80% in radius isolated solar-composition planets of 10 , 1 , and $0.1 M_J$ have $n = 1.3$, 1.0 , and 0.6 ,

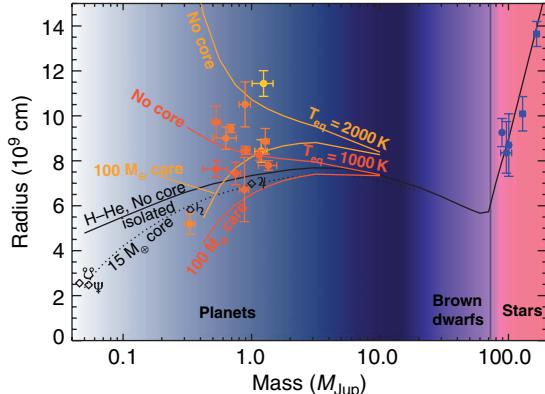


Figure 7 Theoretical and observed mass–radius relations. The black line is applicable to the evolution of solar-composition planets, brown dwarfs and stars, when isolated or nearly isolated (as Jupiter, Saturn, Uranus, and Neptune, defined by diamonds and their respective symbols), after 5 Ga of evolution. The dotted line shows the effect of a $15 M_{\oplus}$ core on the mass–radius relation. Orange and yellow curves represent the mass–radius relations for heavily irradiated planets with equilibrium temperatures of 1000 and 2000 K, respectively, and assuming that 0.5% of the incoming stellar luminosity is dissipated at the center (see Section 10.13.4.3). For each irradiation level, two cases are considered: a solar-composition planet with no core (top curve), and one with a $100 M_{\oplus}$ central core (bottom curve). The transiting extrasolar giant planets for which a mass and a radius was measured are shown with points that are color-coded in function of the planet's equilibrium temperature. The masses and radii of very low-mass stars are also indicated as blue points with error bars.

respectively. From polytropic equations (e.g., Chandrasekhar, 1939):

$$R \propto K^{n/(3-n)} M^{(1-n)/(3-n)} \quad [13]$$

Assuming that K is independent of mass, one gets $R \propto M^{0.16}$, M^0 , and $M^{-0.18}$ for $M = 10$, 1 , and $0.1 M_{Jup}$, respectively, in relatively good agreement with Figure 7 (the small discrepancies are due to the fact that the intrinsic luminosity and hence K depend on the mass considered).

Figure 7 shows already that the planets in our solar system are not made of pure hydrogen and helium and require an additional fraction of heavy elements in their interior, either in the form of a core, or distributed in the envelope (dotted line).

For extrasolar planets, the situation is complicated by the fact that the intense irradiation that they receive plays a major role in their evolution. The present sample is already quite diverse, with equilibrium temperature (defined as the effective temperature corresponding to the stellar flux

received by the planet) ranging from 1000 to 2500 K. Their composition is also quite variable with some planets having large masses of heavy elements (Sato *et al.*, 2005; Guillot *et al.*, 2006). The orange and yellow curves in Figure 7 show theoretical results for equilibrium temperatures of 1000 and 2000 K, respectively. Two extreme models have been plotted: assuming a purely solar composition planet (top curve), and assuming the presence of a $100 M_{\oplus}$ central core (bottom curve). In each case, an additional energy source proportional to 0.5% of the incoming luminosity was also assumed (see discussion in Section 10.13.4.3 hereafter).

The increase in radius for decreasing planetary mass for irradiated, solar-composition planets with little or no core can be understood using the polytropic relation (eqn [13]), but accounting for variations of K as defined by the atmospheric boundary condition. Using the Eddington approximation, assuming $\kappa \propto P$ and a perfect gas relation in the atmosphere, one can show that $K \propto (M/R^2)^{-1/2n}$ and that therefore $R \propto M^{1/2 - n/(2 - n)}$. With $n=1$, one finds $R \propto M^{-1/2}$. Strongly irradiated hydrogen–helium planets of small masses are hence expected to have the largest radii which qualitatively explain the positions of the extrasolar planets in Figure 7. Note that this estimate implicitly assumes that n is constant throughout the planet. The real situation is more complex because of the growth of a deep radioactive region in most irradiated planets, and because of structural changes between the degenerate interior and the perfect gas atmosphere.

In the case of the presence of a fixed mass of heavy elements, the trend is inverse because of the increase of mean molecular mass (or equivalently core/envelope mass) with decreasing total mass. Thus, small planets with a core are much more tightly bound and less subject to evaporation than those that have no core.

10.13.3.6 Rotation and the Figures of Planets

The mass and radius of a planet informs us on its global composition. Because planets are also rotating, one is allowed to obtain more information on their deep interior structure. The hydrostatic equation becomes more complex, however:

$$\frac{\nabla P}{\rho} = \nabla \left(G \iiint \frac{\rho(\mathbf{r}')}{|\mathbf{r} - \mathbf{r}'|} d^3 r' \right) - \boldsymbol{\Omega} \times (\boldsymbol{\Omega} \times \mathbf{r}) \quad [14]$$

where Ω is the rotation vector. The resolution of eqn [14] is a complex problem. It can however be somewhat simplified by assuming that $|\Omega| \equiv \omega$ is such that the centrifugal force can be derived from a potential. The hydrostatic equilibrium then writes $\nabla P = \rho \nabla U$, and the ‘figure’ of the rotating planet is then defined by the $U = \text{constant}$ level surface.

One can show (e.g., Zharkov and Trubitsyn, 1978) that the hydrostatic equation of a fluid planet can then be written in terms of the mean radius \bar{r} (the radius of a sphere containing the same volume as that enclosed by the considered equipotential surface):

$$\frac{1}{\rho} \frac{\partial P}{\partial \bar{r}} = -\frac{Gm}{\bar{r}^2} + \frac{2}{3} \omega^2 \bar{r} + \frac{GM}{\bar{R}^3} \bar{r} \varphi_\omega \quad [15]$$

where M and \bar{R} are the total mass and mean radius of the planet, and φ_ω is a slowly varying function of \bar{r} . (In the case of Jupiter, φ_ω varies from about 2×10^{-3} at the center to 4×10^{-3} at the surface.) Equations [5]–[7] remain the same with the hypothesis that the level surfaces for the pressure, temperature, and luminosity are equipotentials. The significance of rotation is measured by the ratio of the centrifugal acceleration to the gravity:

$$q = \frac{\omega^2 R_{\text{eq}}^3}{GM} \quad [16]$$

As discussed in Section 10.13.2.2, in some cases, the external gravity field of a planet can be accurately measured in the form of gravitational moments J_k (with zero odd moments for a planet in hydrostatic equilibrium) that measure the departure from spherical symmetry. Together with the mass, this provides a constraint on the interior density profile (see Zharkov and Trubitsyn (1974); see Chapters 10.04, 10.02):

$$M = \iiint \rho(r, \theta) d^3 \tau$$

$$J_{2i} = -\frac{1}{MR_{\text{eq}}^{2i}} \iiint \rho(r, \theta) r^{2i} P_{2i}(\cos \theta) d^3 \tau$$

where $d\tau$ is a volume element and the integrals are performed over the entire volume of the planet.

Figure 8 shows how the different layers inside a planet contribute to the mass and the gravitational moments. The figure applies to Jupiter, but would remain relatively similar for other planets. Note however that in the case of Uranus and Neptune, the core is a sizable fraction of the total planet and contributes both to J_2 and J_4 . Measured gravitational moments thus provide information on the external levels of a

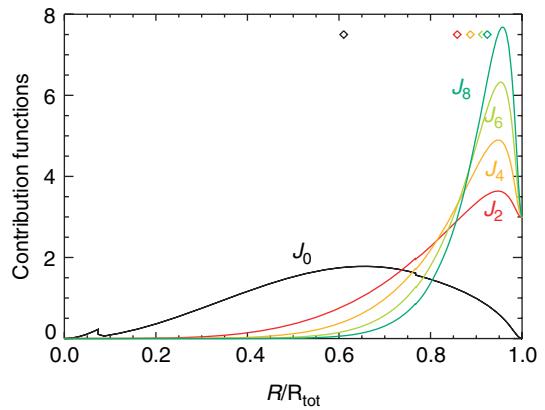


Figure 8 Contribution of the level radii to the gravitational moments of Jupiter. J_0 is equivalent to the planet’s mass. The small discontinuities are caused by the following transitions, from left to right: core/envelope, helium rich/helium poor (metallic/molecular). Diamonds indicate the median radius for each moment.

planet. It is only indirectly, through the constraints on the outer envelope that the presence of a central core can be inferred. As a consequence, it is impossible to determine this core’s state (liquid or solid), structure (differentiated, partially mixed with the envelope) and composition (rock, ice, helium, etc).

For planets outside the solar system, although measuring their gravitational potential is utopic, their oblateness may be reachable with future space transit observations (Seager and Hui, 2002). Since the oblateness e is, to first order, proportional to q :

$$e = \frac{R_{\text{eq}}}{R_{\text{eq}} - R_{\text{pol}}} \approx \left(\frac{3}{2} \Lambda_2 + \frac{1}{2} \right) q \quad [17]$$

(where $\Lambda_2 = J_2/q \approx 0.1$ to 0.2), it may be possible to obtain their rotation rate, or with a rotation measured from another method, a first constraint on their interior structure.

10.13.4 Interior Structures and Evolutions

10.13.4.1 Jupiter and Saturn

As illustrated by **Figure 9**, the simplest interior models of Jupiter and Saturn matching all observational constraints assume the presence of three main layers: (1) an outer hydrogen–helium envelope, whose global composition is that of the deep atmosphere; (2) an inner hydrogen–helium envelope, enriched in helium because the whole planet has to fit the H/He

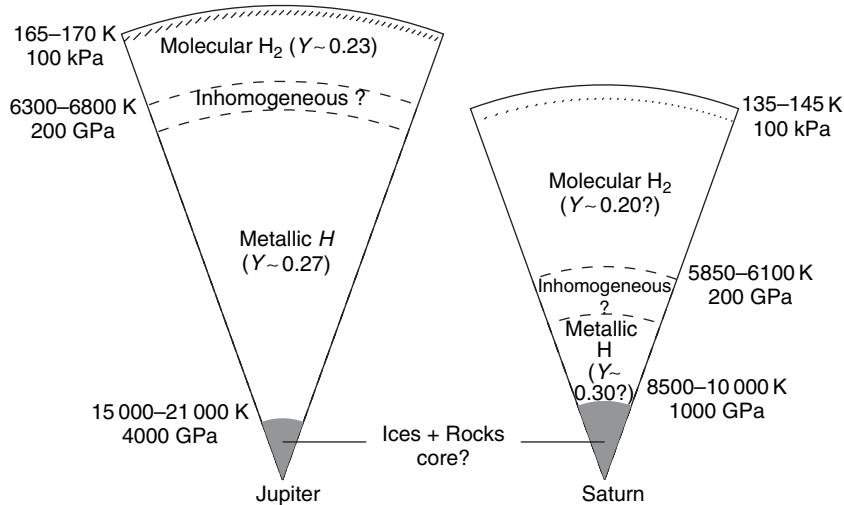


Figure 9 Schematic representation of the interiors of Jupiter and Saturn. The range of temperatures is estimated using homogeneous models and including a possible radiative zone indicated by the hashed regions. Helium mass-mixing ratios Y are indicated. The size of the central rock and ice cores of Jupiter and Saturn is very uncertain (see text). In the case of Saturn, the inhomogeneous region may extend down all the way to the core which would imply the formation of a helium core. Adapted from Guillot T (1999b) Interior of giant planets inside and outside the solar system. *Science* 286: 72–77.

protosolar value; and (3) a central dense core. Because the planets are believed to be mostly convective, these regions are expected to be globally homogeneous. (Many interesting thermochemical transformations take place in the deep atmosphere, but they are of little concern to us.)

The transition from a helium-poor upper envelope to a helium-rich lower envelope is thought to take place through the formation of helium-rich droplets that fall deeper into the planet due to their larger density. These droplets form because of an assumed phase transition of helium in hydrogen at high pressures and low temperatures. Three-layer models implicitly make the hypothesis that this region is narrow. Indeed, calculations of such a phase separation in a fully ionized plasma indicate a rapid decrease of the critical temperature with increasing pressure, with the consequence that helium would be unsoluble in a relatively small region in low-pressure metallic hydrogen. This region would progressively grow with time (e.g., Stevenson, 1982). However, DFT calculations have indicated that the critical temperature for helium demixing may rise with pressure (Pfaffensteller *et al.*, 1995), presumably in the regime where hydrogen is only partially ionized and bound states remain. This opens up the possibility that the inhomogeneous regions may be more extended, and that models more complex than the three-layer models may be needed, in particular in the case of Saturn (see below).

In the absence of these calculations, the three-layer models can be used as a useful guidance to a necessarily hypothetical ensemble of allowed structures and compositions of Jupiter and Saturn. These relatively extensive exploration of the parameter space have been performed by Saumon and Guillot (2004). The calculations assume that only helium is inhomogeneous in the envelope (the abundance of heavy elements is supposed to be uniform across the molecular/metallic hydrogen transition). Many sources of uncertainties are taken into account, however; among them, the most significant are on the EOS of hydrogen and helium, the uncertain values of \mathcal{J}_4 and \mathcal{J}_6 , the presence of differential rotation deep inside the planet, the location of the helium-poor to helium-rich region, and the uncertain helium-to-hydrogen protosolar ratio.

Their results indicate that Jupiter's core is smaller than $\sim 10 M_{\oplus}$, and that its global composition is pretty much unknown (between 10 and $42 M_{\oplus}$ of heavy elements in total). The models indicate that Jupiter is enriched compared to the solar value by a factor 1.5 to 8 times the solar value. This enrichment is compatible with a global uniform enrichment of all species near the atmospheric Galileo values, but include many other possibilities.

In the case of Saturn, the solutions depend less on the hydrogen EOS because the Mbar pressure region is comparatively smaller. The total amount of heavy

elements present in the planet can therefore be estimated with a better accuracy than for Jupiter, and is between 20 and $30 M_{\oplus}$. In three-layer models with a discontinuity of the helium abundance at the molecular–metallic hydrogen interface but continuity of all other elements, the core masses found are between 10 and $22 M_{\oplus}$. However, because Saturn's metallic region is deeper into the planet, it mimics the effect that a central core would have on J_2 . If we allow for variations in the abundance of heavy elements together with the helium discontinuity, then the core mass can become much smaller, and even solutions with no core can be found (Guillot, 1999a). These solutions depend on the hypothetic phase separation of an abundant species (e.g., water), and generally cause an energy problem because of the release of considerable gravitational energy. However, another possibility is through the formation of an almost pure helium shell around the central core, which could lower the core masses by up to $7 M_{\oplus}$ (Fortney and Hubbard, 2003; Hubbard, personal communication).

Concerning the ‘evolutions’ of Jupiter and Saturn, the three main sources of uncertainty are, by order of importance: (1) the magnitude of the helium separation; (2) the EOS; and (3) the atmospheric boundary conditions. Figure 10 shows an ensemble of

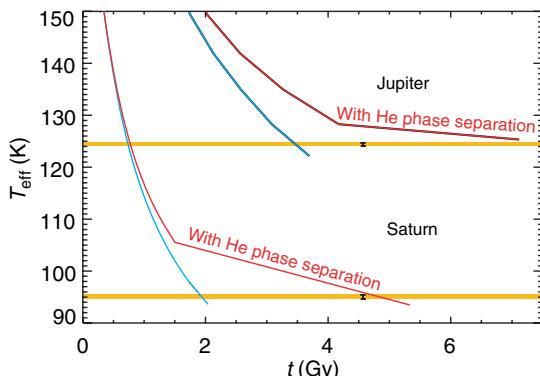


Figure 10 Final stages of evolution of Jupiter and Saturn. The present effective temperatures, reached after ~ 4.55 Ga of evolution, are indicated as horizontal orange lines. For each planet two models represent attempts to bracket the ensemble of possibilities, with the faster evolution corresponding to that of an homogeneous planet, while the slowest evolution includes the effect of helium settling in the last evolution phase. Adapted from Hubbard WB, Guillot T, Marley MS, Burrows A, Lunine JI, and Saumon DS (1999) Comparative evolution of Jupiter and Saturn. *Planetary and Space Science* 47: 1175–1182. Fortney JJ and Hubbard WB (2003) Phase separation in giant planets: Inhomogeneous evolution of Saturn. *Icarus* 164: 228–243.

possibilities that attempt to bracket the minimum and maximum cooling. In all cases, helium sedimentation is needed to explain Saturn's present luminosity (see Salpeter, 1973; Stevenson and Salpeter, 1977a; Hubbard, 1977). Recent models of Saturn's evolution appear to favor a scenario in which helium settles down almost to the central core (Hubbard *et al.*, 1999; Fortney and Hubbard, 2003). In the case of Jupiter, the sedimentation of helium that appears to be necessary to explain the low atmospheric helium abundance poses a problem for evolution models because it appears to generally prolong its evolution beyond 4.55 Ga, the age of the solar system. However, different solutions are possible, including improvements of the EOS and atmospheric boundary conditions, or even the possible progressive erosion of the central core that would yield a lower Jupiter's luminosity at a given age (Guillot *et al.*, 2004).

10.13.4.2 Uranus and Neptune

Although the two planets are relatively similar, Figure 7 already shows that Neptune's larger mean density compared to Uranus has to be due to a slightly different composition: either more heavy elements compared to hydrogen and helium, or a larger rock/ice ratio. The gravitational moments impose that the density profiles lie close to that of ‘ices’ (a mixture initially composed of e.g., H_2O , CH_4 , and NH_3 , but which rapidly becomes a ionic fluid of uncertain chemical composition in the planetary interior), except in the outermost layers, which have a density closer to that of hydrogen and helium (Marley *et al.*, 1995; Podolak *et al.*, 2000). As illustrated in Figure 11, three-layer models of Uranus and Neptune consisting of a central ‘rocks’ core (magnesium-silicate and iron material), an ice layer, and a hydrogen–helium gas envelope have been calculated (Podolak *et al.*, 1991; Hubbard *et al.*, 1995).

The fact that models of Uranus assuming homogeneity of each layer and adiabatic temperature profiles fail in reproducing its gravitational moments seem to imply that substantial parts of the planetary interior are not homogeneously mixed (Podolak *et al.*, 1995). This could explain the fact that Uranus' heat flux is so small: its heat would not be allowed to escape to space by convection, but through a much slower diffusive process in the regions of high molecular-weight gradient. Such regions would also be present in Neptune, but much deeper, thus allowing more heat to be transported outward. The existence

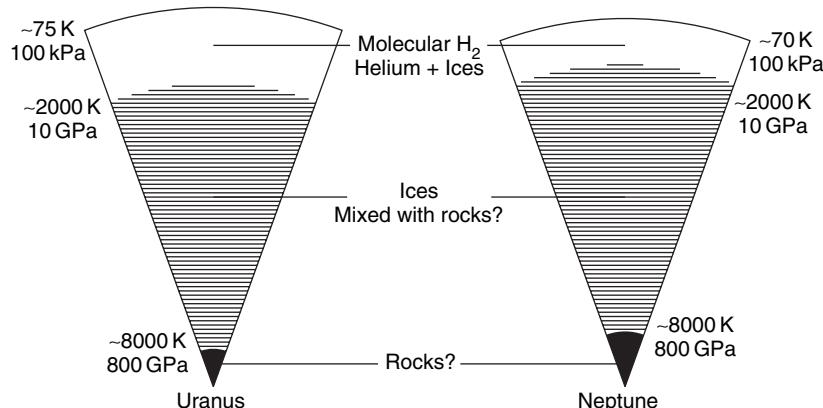


Figure 11 Schematic representation of the interiors of Uranus and Neptune. Adapted from Guillot T (1999b) Interior of giant planets inside and outside the solar system. *Science* 286: 72–77.

of these nonhomogeneous, partially mixed regions are further confirmed by the fact that if hydrogen is supposed to be confined solely to the hydrogen–helium envelope, models predict ice/rock ratios of the order of 10 or more, much larger than the protosolar value of ~ 2.5 . On the other hand, if we impose the constraint that the ice/rock ratio is protosolar, the overall composition of both Uranus and Neptune is, by mass, about 25% rock, 60–70% ice, and 5–15% hydrogen and helium (Podolak *et al.*, 1991, 1995; Hubbard *et al.*, 1995). Assuming both ice and rock are present in the envelope, an upper limit to the amount of hydrogen and helium present is $\sim 4.2 M_{\oplus}$ for Uranus and $\sim 3.2 M_{\oplus}$ for Neptune (Podolak *et al.*, 2000). A lower limit of $\sim 0.5 M_{\oplus}$ for both planets can be inferred by assuming that hydrogen and helium are only present in the outer envelope at $P \lesssim 100$ kbar.

10.13.4.3 Irradiated Giant Planets

Although all extrasolar giant planets are in principle interesting, we focus here on the ones that orbit extremely close to their star because of the possibility to directly characterize them and measure their mass, radius, and some properties of their atmosphere. Two planets are proxies for this new class of objects: the first extrasolar giant planet discovered, 51 Peg b, with an orbital period of $P = 4.23$ days, and the first ‘transiting’ extrasolar giant planet, HD 209458 b, with $P = 3.52$ days. Both planets belong to the Pegasus constellation, and following astronomical conventions (e.g., Cepheids, named after δ Cephei), we choose to name giant planets orbiting close to their stars with periods shorter than 10 days ‘Pegasids’ (alternatively, ‘hot Jupiters’ is also found in the literature).

With such a short orbital period, these planets are for most of them subject to an irradiation from their central star that is so intense that the absorbed stellar energy flux can be about $\sim 10^4$ times larger than their intrinsic flux. The atmosphere is thus prevented from cooling, with the consequence that a radiative zone develops and governs the cooling and contraction of the interior (Guillot *et al.*, 1996). Typically, for a planet like HD 209458 b, this radiative zone extends to kilobar levels, $T \sim 4000$ K, and is located in the outer 5% in radius (0.3% in mass) (Guillot and Showman, 2002).

Problems in the modeling of the evolution of Pegasids arise mostly because of the uncertain outer boundary condition. The intense stellar flux implies that the atmospheric temperature profile is extremely dependent upon the opacity sources considered. Depending on the chosen composition, the opacity data used, the assumed presence of clouds, the geometry considered, resulting temperatures in the deep atmosphere can differ by up to ~ 600 K (Seager and Sasselov, 2000; Goukenleuque *et al.*, 2000; Barman *et al.*, 2001; Sudarsky *et al.*, 2003; Iro *et al.*, 2005; Fortney *et al.*, 2006). Furthermore, as illustrated by Figure 12, the strong irradiation and expected synchronisation of the planets implies that strong inhomogeneities should exist in the atmosphere within particular strong (~ 500 K) day-night and equator-to-pole differences in effective temperatures (Showman and Guillot, 2002; Iro *et al.*, 2005; Cooper and Showman, 2005; Barman *et al.*, 2005), further complicating the modeling of the planetary evolution (see Figure 13). Finally, another related problem is the presence of the radiative zone. Again, the composition is unknown and the opacity data are

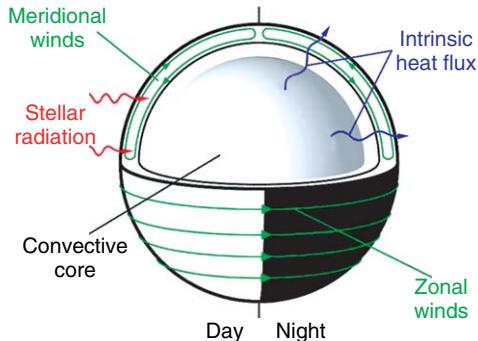


Figure 12 Conjectured dynamical structure of Pegasids (strongly irradiated extrasolar giant planets): at pressures larger than 100–800 bar, the intrinsic heat flux must be transported by convection. The convective core is at or near synchronous rotation with the star and has small latitudinal and longitudinal temperature variations. At lower pressures a radiative envelope is present. The top part of the atmosphere is penetrated by the stellar light on the day side. The spatial variation in insolation should drive winds that transport heat from the day side to the night side. From Showman AP and Guillot T (2002) Atmospheric circulation and tides of ‘51 Pegasus b-like’ planets. *Astronomy and Astrophysics* 385: 166–180.

uncertain in this relatively high temperature ($T \sim 1500\text{--}3000\text{ K}$) and high pressure (up to $\sim 1\text{ kbar}$) regime.

We have seen in [Figure 7](#) that the measured masses and radii of transiting planets can be globally explained in the framework of an evolution model including the strong stellar irradiation and the presence of a variable mass of heavy elements, either in the form of a central core, or spread in the planet interior. However, when analyzing the situation for each planet, it appears that several planets are too large to be reproduced by standard models, that is, models using the most up-to-date EOS, opacities, atmospheric boundary conditions, and assuming that the planetary luminosity governing its cooling is taken solely from the lost gravitational potential energy (see [Section 10.13.3.1](#)).

[Figure 14](#) illustrates the situation for the particular case of HD209458b: unless using an unrealistically hot atmosphere, or arbitrarily increasing the internal opacity, or decreasing the helium content, one cannot reproduce the observed radius which is 10–20% larger than calculated ([Bodenheimer et al., 2001, 2003; Guillot and Showman, 2002; Baraffe et al., 2003](#)). The fact that the measured radius corresponds to a low-pressure ($\sim \text{Mbar}$) level while the calculated radius corresponds to a level near 1 bar is not negligible ([Burrows et al., 2003](#)) but too small to account for

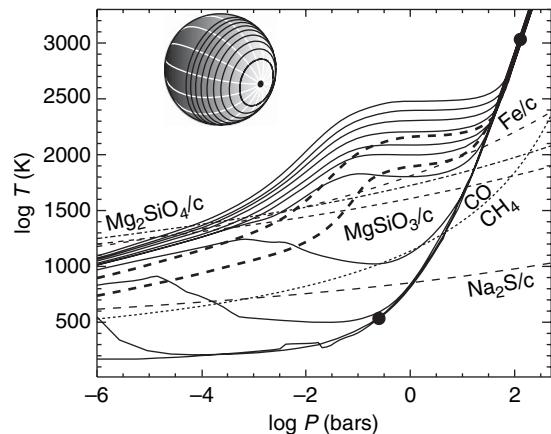


Figure 13 Temperature vs pressure for a sequence of locations in the atmosphere of HD209458b, assuming no horizontal redistribution of heat. Each sequence corresponds to a given direction of the incident flux relative to the surface normal. The approximate regions represented by the collection of T - P profiles are shown as solid black lines on the illustrative sphere. The topmost T - P profile corresponds to the substellar point (black dot on the sphere). The terminator and night side (black hemisphere) are modeled with the nonirradiated profile (lowest T - P curve). The radiative–convective boundary at the substellar point and on the night side are labeled with filled circles. The dashed lines indicate the approximate condensation curves for three common grain species. The dotted line indicates where gaseous CO and CH_4 concentrations are equal (CO is dominant to the left of this line). The thick, gray, dashed lines are T - P profiles calculated for a normal incident flux equal to 0.5 (top) and 0.25 (bottom) times that at the substellar point, as often used as approximate solutions for the day side, or entire atmosphere, respectively. From Barman TS, Hauschildt PH, and Allard F (2005) Phase-dependent properties of extrasolar planet atmospheres. *Astrophysical Journal* 632: 1132–1139.

the difference. This is problematic because while it is easy to invoke the presence of a massive core to explain the small size of a planet, a large size such as that of HD209458b requires an additional energy source, or significant modifications in the data/physics involved.

[Bodenheimer et al. \(2001\)](#) proposed that this large radius may be due to a small forced eccentricity ($e \sim 0.03$) of HD209458b, and subsequent tidal dissipation in the planet interior, but detailed observations indicate that the eccentricity is small, $e = 0.014 \pm 0.009$ ([Laughlin et al., 2005](#)), and observations of the secondary eclipse imply that this would further require a chance configuration of the orbit ([Deming et al., 2005](#)). Another proposed explanation also involving tidal dissipation of orbital energy is that the planet may be trapped in a Cassini state with

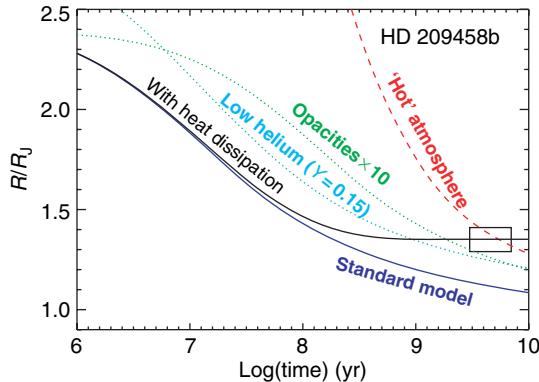


Figure 14 The contraction of HD209458b as a function of time can be compared to its measured radius and inferred age shown by the black box. Standard models (blue curve) for the evolution of that $0.69 M_J$ planet generally yield a radius that is too small compared to the observations, even for a solar composition and no central core (a larger core and, in most cases, larger amounts of heavy elements in the planet imply an even smaller size for a given age). Unrealistically low helium abundances or high opacities models lead to evolution tracks that barely cross the observational box. A possibility is that heat is dissipated into the deep interior by stellar tides, either related to a nonzero orbital eccentricity forced by an unseen companion, or because of a constant transfer of angular momentum from the heated atmosphere to the interior (black curve). Alternatively, the atmosphere may be hotter than predicted due to heating by strong zonal winds and shear instabilities (red curve).

a large orbital inclination (Winn and Holman, 2005), but it appears to have a low probability of occurrence (Levrard *et al.*, 2006). Finally, a third possibility that would apply to ‘all’ Pegasids is to invoke a downward transport of kinetic energy and its dissipation by tides (Showman and Guillot, 2002). This last possibility would require the various transiting planets to have different core masses to reproduce the observed radii (Guillot, 2005).

Recently, as more transiting Pegasids have been discovered, the number of anomalously large ones has increased to at least 3 for 11 planets, implying that this is not a rare event. This lends more weight to a mechanism that would apply to each planet. In this case, masses of heavy elements can be derived by imposing that all planets should be fitted by the same model with the same hypotheses. This can be done by inverting the results of Figure 7, as described by Guillot *et al.* (2006). The method is applied to the known transiting Pegasids by the end of 2006 in Figure 15, a plot of the masses of heavy elements in the planets as a function of the metallicities of the

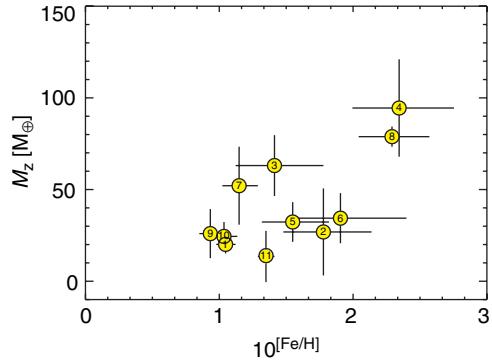


Figure 15 Mass of heavy elements in transiting Pegasids known by 2006 as a function of the metal content of the parent star relative to the Sun. The mass of heavy elements required to fit the measured radii is calculated on the basis of evolution models including an additional heat source slowing the cooling of the planet. This heat source is assumed equal to 0.5% of the incoming stellar heat flux (Showman and Guillot, 2002). Horizontal error bars correspond to the 1σ errors on the $[{\rm Fe/H}]$ determination. Vertical error bars are a consequence of the uncertainties on the measured planetary radii and ages. Note that the results, based on Guillot *et al.* (2006) are intrinsically model-dependent and may be affected by further discoveries of transiting planets.

parent star (which measures how rich a given star is in heavy elements compared to the Sun).

Figure 15 first shows that in some cases, large masses of heavy elements (up to $\sim 100 M_{\oplus}$) are necessary. This is in harmony with the composition inferred for HD149026b, that is, around $70 M_{\oplus}$ of heavy elements, a conclusion that is hard to escape because of the low total mass and high irradiation of the planet (see Ikoma *et al.*, 2006; Fortney *et al.*, 2006). Furthermore, there seems to be a correlation between the mass of heavy elements inferred in the transiting planets, and the metallicity of the parent stars (Guillot *et al.*, 2006), although this correlation has to be ascertained by more measurements. A caveat is important: these results are intrinsically model-dependent as they are based on the assumption that all planets receive an additional tidal heat flux that is proportional to the energy flux that they receive in the form of photons from the star. More transiting planets are needed to confirm or infirm this model, but the large variety in core masses, and the absence of Pegasids around metal-poor stars ($[{\rm Fe/H}] \lesssim -0.07$) as indicated by Figure 15 appear to be robust consequences of this work.

Another intriguing possibility concerning Pegasids is that of a sustained mass loss due to the

high irradiation dose that the planets receive. Indeed, this effect was predicted (Burrows and Lunine, 1995; Guillot *et al.*, 1996; Lammer *et al.*, 2003) and detected (Vidal-Madjar *et al.*, 2003, 2004), but its magnitude is still quite uncertain, by at least two orders of magnitude (Lammer *et al.*, 2003; Lecavelier des Etangs *et al.*, 2004; Yelle, 2006). The effect on the evolution is surprisingly limited, except at the final stages when an exponential mass loss appears in fully gaseous planets (Baraffe *et al.*, 2004).

Finally, it is important to note that another class of planets awaits a direct characterization by the transit method: that of ice or rock giants. Small-mass planets around $10 M_{\oplus}$ have been detected (e.g., Lovis *et al.*, 2006; Beaulieu *et al.*, 2006) but their radius is expected to be small (Guillot *et al.*, 1996; Valencia *et al.*, 2006), and we currently may not have the observational capability to test whether they transit in front of their star. This should be resolved by the space mission CoRoT (launched on 27 Dec 2006) and Kepler (launch \sim 2008). These objects are especially interesting but pose difficult problems in terms of structure because depending on their formation history, precise composition, and location, they may be fluid, solid, or they may even possess a global liquid ocean (see Kuchner, 2003; Léger *et al.*, 2004).

10.13.5 Implications for Planetary Formation Models

The giant planets in our solar system have in common possessing a large mass of hydrogen and helium, but they are obviously quite different in their aspect and in their internal structures. Although studies cannot be conducted with the same level of details, we can safely conclude that extrasolar planets show an even greater variety in composition and visible appearance.

A parallel study of the structures of our giant planets and of giant planets orbiting around other stars should provide us with key information regarding planet formation in the next decade or so. But, already some conclusions, some of them robust, others still tentative, can be drawn (see Chapter 9.01):

1. *Giant planets formed in circumstellar disks, before these were completely dissipated:* This is a relatively obvious consequence of the fact that giant planets are mostly made of hydrogen and helium: these elements had to be acquired when they were still present in the disk. Because the observed lifetime

of gaseous circumstellar disks is of the order of a few million years, this implies that these planets formed (i.e., acquired most of their final masses) in a few million years also, quite faster than terrestrial planets in the solar system.

2. *Giant planets migrated:* Although not cleanly demonstrated yet, there is evidence that the observed orbital distribution of extrasolar planets requires an inward migration of planets, and various mechanisms have been proposed for that (see Ida and Lin, 2004a; Alibert *et al.*, 2005; Moorhead and Adams, 2005, etc.). Separately, it was shown that several properties of our solar system can be explained if Jupiter, Saturn, Uranus, and Neptune ended up the early formation phase in the presence of a disk with quasi-circular orbit, and with Saturn, Uranus, and Neptune significantly closer to the Sun than they are now, and that these three planets subsequently migrated outward (Tsiganis *et al.*, 2005).
3. *Accretion played a key role for giant planet formation:* Several indications point towards a formation of giant planets that is dominated by accretion of heavy elements: first, Jupiter, Saturn, Uranus, and Neptune are all significantly enriched in heavy elements compared to the Sun. This feature can be reproduced by core-accretion models, for Jupiter and Saturn at least (Alibert *et al.*, 2005). Second, the probability to find a giant planet around a solar-type star (with stellar-type F, G, or K) is a strongly rising function of stellar metallicity (Gonzalez, 1998; Santos *et al.*, 2004; Fischer and Valenti, 2005), a property that is also well-reproduced by standard core-accretion models (Ida and Lin, 2004b; Alibert *et al.*, 2005). Third, the large masses of heavy elements inferred in some transiting extrasolar planets as well as the apparent correlation between mass of heavy elements in the planet and stellar metallicity (Guillot *et al.* (2006); see also Sato *et al.* (2005) and Ikoma *et al.* (2006)) is a strong indication that accretion was possible and that it was furthermore efficient. It is to be noted that none of these key properties are directly explained by formation models that assume a direct gravitational collapse (see Boss, 2004; Mayer *et al.*, 2004).
4. *Giant planets were enriched in heavy elements by core accretion, planetesimal delivery and/or formation in an enriched protoplanetary disk:* The giant planets in our solar system are unambiguously enriched in heavy elements compared to the Sun, both globally, and when considering their atmosphere.

This may also be the case of extrasolar planets, although the evidence is still tenuous. The accretion of a central core can explain part of the global enrichment, but not that of the atmosphere. The accretion of planetesimals may be a possible solution but in the case of Jupiter at least the rapid drop in accretion efficiency as the planet reaches appreciable masses ($\sim 100 M_{\oplus}$ or so) implies that such an enrichment would have originally concerned only very deep layers, and would require a relatively efficient upper mixing of these elements, and possibly an erosion of the central core (Guillot *et al.*, 2004).

Although not unambiguously explained, the fact that Jupiter is also enriched in noble gases compared to the Sun is a key observation to understand some of the processes occurring in the early solar system. Indeed, noble gases are trapped into solids only at very low temperatures, and this tells us either that most of the solids that formed Jupiter were formed at very low temperature to be able to trap gases such as argon, probably as clathrates (Gautier *et al.*, 2001; Hersant *et al.*, 2004), or that the planet formed in an enriched disk as it was being evaporated (Guillot and Hueso, 2006).

10.13.6 Future Prospects

We have shown that the compositions and structures of giant planets remain very uncertain. This is an important problem when attempting to understand and constrain the formation of planets, and the origins of the solar system. However, the parallel study of giant planets in our solar system by space missions such as Galileo and Cassini, and of extrasolar planets by both ground-based and space programs has led to rapid improvements in the field, with in particular a precise determination of the composition of Jupiter's troposphere, and constraints on the compositions of a dozen of extrasolar planets (see Chapters 10.17 and 10.18 for discussions of planetary exploration missions).

Improvements on our knowledge of the giant planets requires a variety of efforts. Fortunately, nearly all of these are addressed at least partially by adequate projects in the next few years. The efforts that are necessary thus include (but are not limited to):

- Obtain a better EOS of hydrogen, in particular near the molecular/metallic transition. This will be addressed by the construction of powerful lasers

such as the NIF in the US and the MégaJoule laser in France, and by innovative experiments such as shocks on precompressed samples. One of the challenges is not only obtaining higher pressures, but mostly lower temperatures than currently possible with single shocks. The parallel improvement of computing facilities should allow more extended numerical experiments.

- Calculate hydrogen–helium and hydrogen–water phase diagrams. (Other phase diagrams are desirable too, but of lesser immediate importance.) This should be possible with new numerical experiments.
- Have a better yardstick to measure solar and protosolar compositions. This may be addressed by the analysis of the Genesis mission samples, or may require another future mission.
- Improve the values of \mathcal{J}_4 and \mathcal{J}_6 for Saturn. This will be done as part of the Cassini–Huygens mission. This should lead to better constraints, and possibly a determination of whether the interior of Saturn rotates as a solid body.
- Detect new transiting extrasolar planets, and hopefully some that are further from their star. The space missions CoRoT (2006) and Kepler (2008) should provide the detection and characterization of many tens, possibly hundreds of giant planets.
- Model the formation and evolution of ice giants such as Uranus, Neptune, and similar planets around other stars, in order to analyze detections of these objects and understand planetary formation.
- Improve the measurement of Jupiter's gravity field, and determine the abundance of water in the deep atmosphere. This will be done by the Juno mission (launch 2011) with a combination of an exquisite determination of the planet's gravity field and of radiometric measurements to probe the deep water abundance.
- It would be highly desirable to send a probe similar to the Galileo probe into Saturn's atmosphere. The comparison of the abundance of noble gases would discriminate between different models of the enrichment of the giant planets, and the additional measurement of key isotopic ratio would provide further tests to understand our origins.

Clearly, there is a lot of work on the road, but the prospects for a much improved knowledge of giant planets and their formation are bright.

References

- Acuna MH, Connerney JEP, and Ness NF (1983) The Z3 zonal harmonic model of Saturn's magnetic field analyses and implications. *Journal of Geophysical Research* 88: 8771–8778.
- Alibert Y, Mordasini C, Benz W, and Winisdoerffer C (2005) Models of giant planet formation with migration and disc evolution. *Astronomy and Astrophysics* 434: 343–353.
- Anderson JD, Campbell JK, Jacobson RA, Sweetnam DN, and Taylor AH (1987) Radio science with Voyager 2 at Uranus – Results on masses and densities of the planet and five principal satellites. *Journal of Geophysical Research* 92: 14877–14883.
- Atreya SK, Wong MH, Owen TC, et al. (1999) A comparison of the atmospheres of Jupiter and Saturn: Deep atmospheric composition, cloud structure, vertical mixing, and origin. *Planetary and Space Science* 47: 1243–1262.
- Bahcall JN, Pinsonneault MH, and Wasserburg GJ (1995) Solar models with helium and heavy-element diffusion. *Reviews of Modern Physics* 67: 781–808.
- Baraffe I, Chabrier G, Barman TS, Allard F, and Hauschildt PH (2003) Evolutionary models for cool brown dwarfs and extrasolar giant planets. The case of HD 209458. *Astronomy and Astrophysics* 402: 701–712.
- Baraffe I, Chabrier G, Barman TS, Selsis F, Allard F, and Hauschildt PH (2005) Hot-Jupiters and hot-Neptunes: A common origin? *Astronomy and Astrophysics* 436: L47–L51.
- Baraffe I, Selsis F, Chabrier G, et al. (2004) The effect of evaporation on the evolution of close-in giant planets. *Astronomy and Astrophysics* 419: L13–L16.
- Barman TS, Hauschildt PH, and Allard F (2001) Irradiated planets. *Astrophysical Journal* 556: 885–895.
- Barman TS, Hauschildt PH, and Allard F (2005) Phase-dependent properties of extrasolar planet atmospheres. *Astrophysical Journal* 632: 1132–1139.
- Beaulieu J-P, Bennett DP, Fouqué P, et al. (2006) Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing. *Nature* 439: 437–440.
- Belov SI, Boriskov GV, Bykov AI, et al. (2002) Shock compression of solid deuterium. *Journal of Experimental and Theoretical Physics Letters* 76: 433.
- Bézard B, Lellouch E, Strobel D, Maillard J-P, and Drossart P (2002) Carbon monoxide on Jupiter: Evidence for both internal and external sources. *Icarus* 159: 95–111.
- Bodenheimer P, Laughlin G, and Lin DNC (2003) On the radii of extrasolar giant planets. *Astrophysical Journal* 592: 555–563.
- Bodenheimer P, Lin DNC, and Mardling RA (2001) On the tidal inflation of short-period extrasolar planets. *Astrophysical Journal* 548: 466–472.
- Boncza SA, Militzer B, and Galli G (2004) *Ab initio* simulations of dense liquid deuterium: Comparison with gas-gun shock-wave experiments. *Physical Review B* 69(1): 014101.
- Boriskov GV, Bykov AI, Il'Kaev RI, et al. (2005) Shock compression of liquid deuterium up to 109 GPa. *Physical Review B* 71(9): 092104.
- Borysow A, Jorgensen UG, and Zheng C (1997) Model atmospheres of cool, low-metallicity stars: The importance of collision-induced absorption. *Astronomy and Astrophysics* 324: 185–195.
- Boss AP (2004) Convective cooling of protoplanetary disks and rapid giant planet formation. *Astrophysical Journal* 610: 456–463.
- Briggs FH and Sackett PD (1989) Radio observations of Saturn as a probe of its atmosphere and cloud structure. *Icarus* 80: 77–103.
- Burrows A, Guillot T, Hubbard WB, et al. (2000) On the radii of close-in giant planets. *Astrophysical Journal Letters* 534: L97–L100.
- Burrows A and Lunine J (1995) Extrasolar planets – Astronomical questions of origin and survival. *Nature* 378: 333.
- Burrows A, Marley M, Hubbard WB, et al. (1997) A nongray theory of extrasolar giant planets and brown dwarfs. *Astrophysical Journal* 491: 856–875.
- Burrows A, Marley MS, and Sharp CM (2000) The near-infrared and optical spectra of methane dwarfs and brown dwarfs. *Astrophysical Journal* 531: 438–446.
- Burrows A, Sudarsky D, and Hubbard WB (2003) A theory for the radius of the transiting giant planet HD 209458b. *Astrophysical Journal* 594: 545–551.
- Busse FH (1978) Magnetohydrodynamics of the Earth's dynamo. *Annual Review of Fluid Mechanics* 10: 435–462.
- Campbell JK and Synott SP (1985) Gravity field of the Jovian system from Pioneer and Voyager tracking data. *Astronomical Journal* 90: 364–372.
- Cavazzoni C, Chiarotti GL, Scandolo S, Tosatti E, Bernasconi M, and Parrinello M (1999) Superionic and metallic states of water and ammonia at giant planet conditions. *Science* 283: 44–46.
- Cecconi B and Zarka P (2005) Model of a variable radio period for Saturn. *Journal of Geophysical Research* 110: A12203.
- Chandrasekhar S (1939) *An Introduction to the Study of Stellar Structure*. Chicago, IL: The University of Chicago Press.
- Charbonneau D, Allen LE, Megeath ST, et al. (2005) Detection of thermal emission from an extrasolar planet. *Astrophysical Journal* 626: 523–529.
- Charbonneau D, Brown TM, Noyes RW, and Gilliland RL (2002) Detection of an extrasolar planet atmosphere. *Astrophysical Journal* 568: 377–384.
- Cohen ER and Taylor BN (1987) The 1986 adjustment of the fundamental physical constants. *Reviews of Modern Physics* 59: 1121–1148.
- Collins GW, da Silva LB, Celliers P, et al. (1998) Measurements of the equation of state of deuterium at the fluid insulator-metal transition. *Science* 281: 1178–1181.
- Connerney JEP, Ness NF, and Acuna MH (1982) Zonal harmonic model of Saturn's magnetic field from Voyager 1 and 2 observations. *Nature* 298: 44–46.
- Conrath BJ and Gautier D (2000) Saturn helium abundance: A reanalysis of Voyager measurements. *Icarus* 144: 124–134.
- Cooper CS and Showman AP (2005) Dynamic meteorology at the photosphere of HD 209458b. *Astrophysical Journal Letters* 629: L45–L48.
- da Silva LB, Celliers P, Collins GW, et al. (1997) Absolute equation of state measurements on shocked liquid deuterium up to 200 GPa (2 Mbar). *Physical Review Letters* 78: 483–486.
- Datchi F, Loubeyre P, and Letolle R (2000) Extended and accurate determination of the melting curves of argon, helium, ice (H_2O), and hydrogen (H_2). *Physical Review B* 61: 6535–6546.
- Davies ME, Abalakin VK, Bursa M, Lederle T, and Lieske JH (1986) Report of the IAU/IAG/COSPAR working group on cartographic coordinates and rotational elements of the planets and satellites – 1985. *Celestial Mechanics* 39: 103–113.
- de Pater I, Romani PN, and Atreya SK (1991) Possible microwave absorption by H_2S gas in Uranus' and Neptune's atmospheres. *Icarus* 91: 220–233.
- Deming D, Seager S, Richardson LJ, and Harrington J (2005) Infrared radiation from an extra-solar planet. *Nature* 434: 740–743.
- Desjarlais MP (2003) Density-functional calculations of the liquid deuterium Hugoniot, reshock, and reverberation timing. *Physical Review B* 68(6): 064204.
- Fegley BJ and Lodders K (1994) Chemical models of the deep atmospheres of Jupiter and Saturn. *Icarus* 110: 117–154.

- Ferguson JW, Alexander DR, Allard F, et al. (2005) Low-temperature opacities. *Astrophysical Journal* 623: 585–596.
- Fischer DA and Valenti J (2005) The planet-metallicity correlation. *Astrophysical Journal* 622: 1102–1117.
- Flasar FM, Achterberg RK, Conrath BJ, et al. (2005) Temperatures, winds, and composition in the Saturnian system. *Science* 307: 1247–1251.
- Fortney JJ and Hubbard WB (2003) Phase separation in giant planets: Inhomogeneous evolution of Saturn. *Icarus* 164: 228–243.
- Fortney JJ, Saumon D, Marley MS, Lodders K, and Freedman RS (2006) Atmosphere, interior, and evolution of the metal-rich transiting planet HD 149026b. *Astrophysical Journal* 642: 495–504.
- Galopeau PHM and Lecacheux A (2000) Variations of Saturn's radio rotation period measured at kilometer wavelengths. *Journal of Geophysical Research* 105: 13089–13102.
- Gautier D, Conrath BJ, Owen T, De Pater I, and Atreya SK (1995) The troposphere of Neptune. In: Cruikshank D and Matthews MS (eds.) *Neptune and Triton*, pp. 547–611. Tucson, AZ: University of Arizona Press.
- Gautier D, Hersant F, Mousis O, and Lunine JI (2001) Erratum: Enrichments in volatiles in Jupiter: A new interpretation of the Galileo measurements. *Astrophysical Journal Letters* 559: L183–L183.
- Giampieri G, Dougherty MK, Smith EJ, and Russell CT (2006) A regular period for Saturn's magnetic field that may track its internal rotation. *Nature* 441: 62–64.
- Gierasch PJ, Ingersoll AP, Banfield D, et al. (2000) Observation of moist convection in Jupiter's atmosphere. *Nature* 403: 628–630.
- Gonzalez G (1998) Spectroscopic analyses of the parent stars of extrasolar planetary system candidates. *Astronomy and Astrophysics* 334: 221–238.
- Goukenleuque C, Bézard B, Joguet B, Lellouch E, and Freedman R (2000) A radiative equilibrium model of 51 Peg b. *Icarus* 143: 308–323.
- Gregoryan E, Goncharov AF, Matsuishi K, Mao H-K, and Hemley RJ (2003) Raman spectroscopy of hot dense hydrogen. *Physical Review Letters* 90(17): 175701.
- Guillot T (1999a) A comparison of the interiors of Jupiter and Saturn. *Planetary and Space Science* 47: 1183–1200.
- Guillot T (1999b) Interior of giant planets inside and outside the solar system. *Science* 286: 72–77.
- Guillot T (2005) The interiors of giant planets: Models and outstanding questions. *Annual Review of Earth and Planetary Sciences* 33: 493–530.
- Guillot T, Burrows A, Hubbard WB, Lunine JI, and Saumon D (1996) Giant planets at small orbital distances. *Astrophysical Journal Letters* 459: L35–L39.
- Guillot T and Hueso R (2006) The composition of Jupiter: Sign of a (relatively) late formation in a chemically evolved protosolar disc. *Monthly Notices of the Royal Astronomical Society* 367: L47–L51.
- Guillot T, Santos NC, Pont F, Iro N, Melo C, and Ribas I (2006) A correlation between the heavy element content of transiting extrasolar planets and the metallicity of their parent stars. *Astronomy and Astrophysics* 453: L21–L24.
- Guillot T and Showman AP (2002) Evolution of '51 pegasus b-like' planets. *Astronomy and Astrophysics* 385: 156–165.
- Guillot T, Stevenson DJ, Hubbard WB, and Saumon D (2004) The interior of Jupiter. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter: The Planet, Satellites and Magnetosphere*, pp. 35–57. Cambridge UK: Cambridge University Press.
- Gulkis S, Janssen MA, and Olsen ET (1978) Evidence for the depletion of ammonia in the Uranus atmosphere. *Icarus* 34: 10–19.
- Gurnett DA, Kurth WS, Hospodarsky GB, et al. (2005) Radio and plasma wave observations at Saturn from Cassini's approach and first orbit. *Science* 307: 1255–1259.
- Hammel HB, de Pater I, Gibbard SG, Lockwood GW, and Rages K (2005) New cloud activity of Uranus in 2004: First detection of a southern feature at 2.2 μm . *Icarus* 175: 284–288.
- Harrington J, Hansen BM, Luszcz SH, et al. (2006) The phase-dependent infrared brightness of the extrasolar planet ν Andromedae b. *Science* 314: 623–626.
- Hersant F, Gautier D, and Lunine JI (2004) Enrichment in volatiles in the giant planets of the solar system. *Planetary and Space Science* 52: 623–641.
- Holmes NC, Ross M, and Nellis WJ (1995) Temperature measurements and dissociation of shock-compressed liquid deuterium and hydrogen. *Physical Review B* 52: 15835–15845.
- Hubbard WB (1968) Thermal structure of Jupiter. *Astrophysical Journal* 152: 745–754.
- Hubbard WB (1977) The Jovian surface condition and cooling rate. *Icarus* 30: 305–310.
- Hubbard WB (1989) Structure and composition of giant planet interiors. In: Atreya SK, Pollack JB, and Matthews MS (eds.) *Origin and Evolution of Planetary and Satellite Atmospheres*, pp. 539–563. Tucson, AZ: University of Arizona Press.
- Hubbard WB, Burrows A, and Lunine JI (2002) Theory of giant planets. *Annual Review of Astronomy and Astrophysics* 40: 103–136.
- Hubbard WB, Guillot T, Marley MS, Burrows A, Lunine JI, and Saumon DS (1999) Comparative evolution of Jupiter and Saturn. *Planetary and Space Science* 47: 1175–1182.
- Hubbard WB, Pearl JC, Podolak M, and Stevenson DJ (1995) The interior of Neptune. In: Cruikshank D and Matthews MS (eds.) *Neptune and Triton*, pp. 109–138. Tucson, AZ: University of Arizona Press.
- Hueso R, Sánchez-Lavega A, and Guillot T (2002) A model for large-scale convective storms in Jupiter. *Journal of Geophysical Research (Planets)* 107: 5–1.
- Ida S and Lin DNC (2004a) Toward a deterministic model of planetary formation. I: A desert in the mass and semimajor axis distributions of extrasolar planets. *Astrophysical Journal* 604: 388–413.
- Ida S and Lin DNC (2004b) Toward a deterministic model of planetary formation. II: The formation and retention of gas giant planets around stars with a range of metallicities. *Astrophysical Journal* 616: 567–572.
- Ikoma M, Guillot T, Genda H, Tanigawa T, and Ida S (2006) On the origin of HD 149026b. *Astrophysical Journal* 650: 1150–1159.
- Ingersoll AP, Barnet CD, Beebe RF, et al. (1995) Dynamic meteorology of Neptune. In: Cruikshank D and Matthews MS (eds.) *Neptune and Triton*, pp. 613–682. Tucson, AZ: University of Arizona Press.
- Iro N, Bézard B, and Guillot T (2005) A time-dependent radiative model of HD 209458b. *Astronomy and Astrophysics* 436: 719–727.
- Jacobson RA, Antreasian PG, Bordi JJ, et al. (2006) The gravity field of the Saturnian system from satellite observations and spacecraft tracking data. *Astronomical Journal* 132: 2520–2526.
- Kippenhahn R and Weigert A (1994) *Stellar Structure and Evolution*, vol. XVI, 468 pp, 192 figs. Berlin, Germany: Springer-Verlag. (Also *Astronomy and Astrophysics Library*).
- Knudson MD, Hanson DL, Bailey JE, Hall CA, Asay JR, and Deeney C (2004) Principal Hugoniot, reverberating wave, and mechanical reheat measurements of liquid deuterium to 400 GPa using plate impact techniques. *Physical Review B* 69(14): 144209.

- Kuchner MJ (2003) Volatile-rich Earth-mass planets in the habitable zone. *Astrophysical Journal Letters* 596: L105–L108.
- Lammer H, Selsis F, Ribas I, Guinan EF, Bauer SJ, and Weiss WW (2003) Atmospheric loss of exoplanets resulting from stellar X-ray and extreme-ultraviolet heating. *Astrophysical Journal Letters* 598: L121–L124.
- Laughlin G, Marcy GW, Vogt SS, Fischer DA, and Butler RP (2005) On the eccentricity of HD 209458b. *Astrophysical Journal Letters* 629: L121–L124.
- Lecavelier des Etangs A, Vidal-Madjar A, McConnell JC, and Hébrard G (2004) Atmospheric escape from hot Jupiters. *Astronomy and Astrophysics* 418: L1–L4.
- Léger A, Selsis F, Sotin C, et al. (2004) A new family of planets? ‘ocean-planets’. *Icarus* 169: 499–504.
- Levrard B, Morgado Correia A, Chabrier G, Baraffe I, Selsis F, and Laskar J (2006) Tidal dissipation within hot Jupiters: A new appraisal. *ArXiv Astrophysics e-prints*.
- Lindal GF (1992) The atmosphere of Neptune – An analysis of radio occultation data acquired with Voyager 2. *Astronomical Journal* 103: 967–982.
- Lindal GF, Sweetnam DN, and Eshleman VR (1985) The atmosphere of Saturn – An analysis of the Voyager radio occultation measurements. *Astronomical Journal* 90: 1136–1146.
- Lindal GF, Wood GE, Levy GS, et al. (1981) The atmosphere of Jupiter – An analysis of the Voyager radio occultation measurements. *Journal of Geophysical Research* 86: 8721–8727.
- Little B, Anger CD, Ingersoll AP, et al. (1999) Galileo images of lightning on Jupiter. *Icarus* 142: 306–323.
- Lodders K and Fegley B, Jr (1994) The origin of carbon monoxide in Neptunes’s atmosphere. *Icarus* 112: 368–375.
- Lois C, Mayor M, Pepe F, et al. (2006) An extrasolar planetary system with three Neptune-mass planets. *Nature* 441: 305–309.
- Mao H-K and Hemley RJ (1994) Ultrahigh-pressure transitions in solid hydrogen. *Reviews of Modern Physics* 66: 671–692.
- Marley MS, Fortney JJ, Hubickyj O, Bodenheimer P, and Lissauer JJ (2007) On the luminosity of young Jupiters. *Astrophysical Journal* 655: 541–549.
- Marley MS, Gómez P, and Podolak M (1995) Monte Carlo interior models for Uranus and Neptune. *Journal of Geophysical Research* 100: 23349–23354.
- Marley MS and McKay CP (1999) Thermal structure of Uranus’ atmosphere. *Icarus* 138: 268–286.
- Mayer L, Quinn T, Wadsley J, and Stadel J (2004) The evolution of gravitationally unstable protoplanetary disks: Fragmentation and possible giant planet formation. *Astrophysical Journal* 609: 1045–1064.
- Mayor M and Queloz D (1995) A Jupiter-mass companion to a solar-type star. *Nature* 378: 355–359.
- Militzer B, Ceperley DM, Kress JD, Johnson JD, Collins LA, and Mazeret S (2001) Calculation of a deuterium double shock Hugoniot from *ab initio* simulations. *Physical Review Letters* 87(26): A265502.
- Moorhead AV and Adams FC (2005) Giant planet migration through the action of disk torques and planet-planet scattering. *Icarus* 178: 517–539.
- Ness NF, Acuna MH, Behannon KW, Burlaga LF, Connerney JEP, and Lepping RP (1986) Magnetic fields at Uranus. *Science* 233: 85–89.
- Ness NF, Acuna MH, Burlaga LF, Connerney JEP, and Lepping RP (1989) Magnetic fields at Neptune. *Science* 246: 1473–1478.
- Owen T, Mahaffy P, Niemann HB, et al. (1999) A low-temperature origin for the planetesimals that formed Jupiter. *Nature* 402: 269–270.
- Pearl JC and Conrath BJ (1991) The albedo, effective temperature, and energy balance of Neptune, as determined from Voyager data. *Journal of Geophysical Research* 96: 18921.
- Pfaffenzeller O, Hohl D, and Ballone P (1995) Miscibility of hydrogen and helium under astrophysical conditions. *Physical Review Letters* 74: 2599–2602.
- Podolak M, Hubbard WB, and Stevenson DJ (1991) Model of Uranus’ interior and magnetic field. In: Bergstrahl J (ed.) *Uranus*, pp. 29–61. Tucson, AZ: University of Arizona Press.
- Podolak M, Podolak JL, and Marley MS (2000) Further investigations of random models of Uranus and Neptune. *Planetary and Space Science* 48: 143–151.
- Podolak M, Weizman A, and Marley M (1995) Comparative models of Uranus and Neptune. *Planetary and Space Science* 43: 1517–1522.
- Rages K, Hammel HB, and Lockwood GW (2002) A prominent apparition of Neptune’s south polar feature. *Icarus* 159: 262–265.
- Ribas I (2006) Masses and radii of low-mass stars: Theory versus observations. *Astrophysics and Space Science* 304: 89–92.
- Ross M (1998) Linear-mixing model for shock-compressed liquid deuterium. *Physical Review B* 58: 669–677.
- Ross M and Yang LH (2001) Effect of chainlike structures on shock-compressed liquid deuterium. *Physical Review B* 64(13): 134210.
- Roulston MS and Stevenson DJ (1995) Prediction of neon depletion in Jupiter’s atmosphere. *EOS* 76: 343.
- Salpeter EE (1973) On convection and gravitational layering in Jupiter and in stars of low mass. *Astrophysical Journal Letters* 181: L83–L86.
- Sánchez-Lavega A, Lecacheux J, Gomez JM, et al. (1996) Large-scale storms in Saturn’s atmosphere during 1994. *Science* 271: 631–634.
- Santos NC, Israelian G, and Mayor M (2004) Spectroscopic [Fe/H] for 98 extra-solar planet-host stars. Exploring the probability of planet formation. *Astronomy and Astrophysics* 415: 1153–1166.
- Sato B, Fischer DA, Henry GW, et al. (2005) The N2K consortium. II: A transiting hot Saturn around HD 149026 with a large dense core. *Astrophysical Journal* 633: 465–473.
- Sauvage D, Chabrier G, and van Horn HM (1995) An equation of state for low-mass stars and giant planets. *Astrophysical Journal Supplement* 99: 713.
- Sauvage D and Guillot T (2004) Shock compression of deuterium and the interiors of Jupiter and Saturn. *Astrophysical Journal* 609: 1170–1180.
- Sauvage D, Hubbard WB, Burrows A, Guillot T, Lunine JI, and Chabrier G (1996) A theory of extrasolar giant planets. *Astrophysical Journal* 460: 993.
- Seager S and Hui L (2002) Constraining the rotation rate of transiting extrasolar planets by oblateness measurements. *Astrophysical Journal* 574: 1004–1010.
- Seager S and Sasselov DD (2000) Theoretical transmission spectra during extrasolar giant planet transits. *Astrophysical Journal* 537: 916–921.
- Seiff A, Kirk DB, Knight TCD, et al. (1998) Thermal structure of Jupiter’s atmosphere near the edge of a 5- μ m hot spot in the north equatorial belt. *Journal of Geophysical Research* 103: 22857–22890.
- Showman AP and Guillot T (2002) Atmospheric circulation and tides of ‘51 Pegasus b-like’ planets. *Astronomy and Astrophysics* 385: 166–180.
- Showman AP and Ingersoll AP (1998) Interpretation of Galileo probe data and implications for Jupiter’s dry downdrafts. *Icarus* 132: 205–220.

- Simon-Miller AA, Gierasch PJ, Beebe RF, et al. (2002) New observational results concerning Jupiter's great red spot. *Icarus* 158: 249–266.
- Stevenson DJ (1982) Interiors of the giant planets. *Annual Review of Earth and Planetary Sciences* 10: 257–295.
- Stevenson DJ (1983) Planetary magnetic fields. *Reports of Progress in Physics* 46: 555–557.
- Stevenson DJ (1985) Cosmochemistry and structure of the giant planets and their satellites. *Icarus* 62: 4–15.
- Stevenson DJ and Salpeter EE (1977a) The dynamics and helium distribution in hydrogen–helium fluid planets. *Astrophysical Journal Supplement* 35: 239–261.
- Stevenson DJ and Salpeter EE (1977b) The phase diagram and transport properties for hydrogen–helium fluid planets. *Astrophysical Journal Supplement* 35: 221–237.
- Sudarsky D, Burrows A, and Hubeny I (2003) Theoretical spectra and atmospheres of extrasolar giant planets. *Astrophysical Journal* 588: 1121–1148.
- Trafton LM (1967) Model atmospheres of the major planets. *Astrophysical Journal* 147: 765–781.
- Tsiganis K, Gomes R, Morbidelli A, and Levison HF (2005) Origin of the orbital architecture of the giant planets of the solar system. *Nature* 435: 459–461.
- Tyler GL, Sweetnam DN, Anderson JD, et al. (1989) Voyager radio science observations of Neptune and Triton. *Science* 246: 1466–1473.
- Valencia D, O'Connell RJ, and Sasselov D (2006) Internal structure of massive terrestrial planets. *Icarus* 181: 545–554.
- Vidal-Madjar A, Désert J-M, Lecavelier des Etangs A, et al. (2004) Detection of oxygen and carbon in the hydrodynamically escaping atmosphere of the extrasolar planet HD 209458b. *Astrophysical Journal Letters* 604: L69–L72.
- Vidal-Madjar A, Lecavelier des Etangs A, Désert J-M, et al. (2003) An extended upper atmosphere around the extrasolar planet HD209458b. *Nature* 422: 143–146.
- von Zahn U, Hunten DM, and Lehmacher G (1998) Helium in Jupiter's atmosphere: Results from the Galileo probe helium interferometer experiment. *Journal of Geophysical Research* 103: 22815–22830.
- Warwick JW, Evans DR, Peltzer GR, et al. (1989) Voyager planetary radio astronomy at Neptune. *Science* 246: 1498–1501.
- Warwick JW, Evans DR, Romig JH, et al. (1986) Voyager 2 radio observations of Uranus. *Science* 233: 102–106.
- Weir ST, Mitchell AC, and Nellis WJ (1996) Metallization of fluid molecular hydrogen at 140 GPa (1.4 Mbar). *Physical Review Letters* 76: 1860–1863.
- Winn JN and Holman MJ (2005) Obliquity tides on hot Jupiters. *Astrophysical Journal Letters* 628: L159–L162.
- Wong MH, Mahaffy PR, Atreya SK, Niemann HB, and Owen TC (2004) Updated Galileo probe mass spectrometer measurements of carbon, oxygen, nitrogen, and sulfur on Jupiter. *Icarus* 171: 153–170.
- Yelle RV (2006) Corrigendum to ‘aeronomy of extra-solar giant planets at small orbital distances’ [Icarus 170 (2004) 167–179]. *Icarus* 183: 508–508.
- Zharkov VN and Trubitsyn VP (1974) Internal constitution and the figures of the giant planets. *Physics of the Earth and Planetary Interiors* 8: 105–107.
- Zharkov VN and Trubitsyn VP (1978) *Astronomy and Astrophysics Series: Physics of Planetary Interiors*. Tucson, AZ: Pachart.

10.14 The Origin of the Natural Satellites

S. J. Peale, University of California, Santa Barbara, CA, USA

© 2007 Elsevier B.V. All rights reserved.

10.14.1	Introduction	465
10.14.2	The Earth–Moon System	467
10.14.3	Mars System	471
10.14.4	Jupiter System	473
10.14.4.1	Minimum Mass Sub-Nebula (MMSN)	475
10.14.4.2	Planetesimal Capture Model	476
10.14.4.3	Starved Accretion Disk Model	476
10.14.5	Saturn System	482
10.14.6	Uranus System	485
10.14.7	Neptune System	487
10.14.8	Pluto System	490
10.14.9	Irregular Satellites	491
	References	503

10.14.1 Introduction

This chapter is a discussion of current thoughts on the origin of the natural satellites, but it is not a review of all mechanisms of origin that have been proposed. Rather the most recent papers will be selected with the assumption that these will usually improve what has been published earlier by incorporating more recent results and arriving hopefully at more plausible models. The degree of completeness and plausibility will necessarily be nonuniform among the systems, where we think we understand the history of some systems reasonably well, but others remain mysterious. Inevitably, the author's own prejudices and judgments will emerge, but these have been modified and massaged and sometimes refuted by the advice from several colleagues, who kindly read and commented on the manuscript. Hopefully this chapter can serve as a beginning for researchers to continue the quest to understand this part of solar system origin.

Of all the planetary satellites, our own Moon has generated the most curiosity about its origin and the subsequent tidal evolution that has placed constraints on the theories. A remarkable number of diverse theories of the origin of the Moon have emerged, and within these we find the mechanisms of origin for the vast number of other planetary satellites (e.g., Hartman *et al.*, 1986; Canup and Righter, 2000). Each system of satellites is unique, and it is

impossible to apply the same detailed sequence of events to account for the origin of the different systems. The dynamical mechanisms that have been proposed for the formation of the Moon are evaluated by Boss and Peale (1986), and one or another of these have been proposed for other natural satellites. The mechanisms include what have been called: rotational fission, precipitation fission, intact capture, disintegrative capture, binary accretion, and formation by giant impact. In rotational fission, favored by Darwin (1879, 1880), the Earth spins so rapidly that it separates into two bodies, the Earth and Moon, as a result of a dynamic fission instability. Precipitation fission (Ringwood, 1970, 1972) requires a massive primordial atmosphere sufficiently hot to vaporize silicates. As the atmosphere cools, the silicates condense and precipitate to the Earth's equatorial plane where they can collect into the Moon, while the volatiles in the atmosphere are swept away by a T-Tauri solar wind. For intact capture, the Moon is assumed to form from the accretion of planetesimals in heliocentric orbit relatively close to that of the Earth. During a close approach to the Earth, energy must be removed from the Moon to leave it in a bound orbit. Methods proposed for this energy loss include collision with an already-existing satellite (Dyczmons, 1978), atmospheric drag (Horedt, 1976), an encounter with another object in heliocentric orbit when both the Moon and the other object are within the Earth's sphere of influence (Ruskol, 1972a),

changes in the masses of the Sun (Szebehely and Evans, 1980) or Earth (Lyttleton, 1967), and the dissipation of tidal energy during the time of close approach (Gerstenkorn, 1955, 1969; Singer, 1968, 1970, 1972; Singer and Bandermann, 1970; Alfvén and Arrhenius, 1969, 1972, 1976; Winters and Malcuit, 1977; Conway 1982). Disintegrative capture involves the tidal disruption of a proto-Moon during a close encounter with the Earth, where some of the resulting fragments could be more easily captured into bound orbit to later accumulate into the Moon (Alfvén, 1963; Alfvén and Arrhenius, 1969; Öpik, 1969, 1971). In binary accretion, the Moon forms in Earth orbit from an accretion disk simultaneously with the formation of the Earth (Ruskol, 1961, 1963, 1972a, 1972b, 1972c, 1975; Harris and Kaula, 1975; Harris, 1977, 1978). A formation by giant impact takes advantage of the fact that the last stages of accretion of the terrestrial planets involved large bodies (Wetherill, 1985). The result of such an impact is likely to leave sufficient material beyond the Roche radius R_R to accumulate into the Moon (Öpik, 1971; Hartmann and Davis, 1975; Cameron and Ward, 1976; Canup, 2004a, 2004b). Only the giant impact origin of the Moon has survived rather intense scrutiny. (See Boss and Peale (1986) for a detailed rejection of the other schemes and Canup (2004a, 2004b) for recent successful attempts to simulate the giant impact.)

As formation of satellites in accretion disks must be the dominant process for the regular satellites of the giant planets (regular satellites are those in nearly circular orbits coplanar with the planet's equator plane), we include a description of the processes and parameters appropriate to such accretion disks in Appendix 1. In Appendix 2 we outline a theory of gravitational tides and the dissipation therein that result in orbital and spin evolution. This evolution in many instances constrains the choice of processes and events involved in the origin of particular satellites. We do not include tables of orbital parameters or physical properties of the satellites, but reader is referred to the websites (see Relevant Websites section) for frequently updated total number of satellites with their orbital parameters at the first website and their updated physical properties in the second.

In Section 10.14.2 we develop the arguments supporting the giant impact origin of the Earth's Moon, which is necessary to account for the large angular momentum of the system along with a volatile and iron poor Moon. The equatorial orbits of Mars' satellites requires their origin in a dissipative disk of

debris, where means of accounting for a suspected grossly different composition of the satellites from Mars itself are discussed in Section 10.14.3. We describe recent attempts at understanding the origin of Jupiter's regular satellites by a generalization of binary accretion during the last stages of Jupiter's formation in Section 10.14.4. The constraints imposed by information obtained from the Galileo spacecraft mission guide these attempts, but all involve assumptions of various degrees of uncertainty. But considerable progress has been made recently, and a self-consistent and plausible theory of origin of Jupiter's regular satellites within the theory of the accretion of Jupiter itself is emerging.

In many ways, Saturn's system of regular satellites is much more complicated than Jupiter's, where nearly all of the mass in the system is in the single large satellite Titan, and small icy satellites appear inside Titan's orbit where we might have expected large rocky satellites as in the Jupiter system. The origin of Saturn's satellites is discussed in Section 10.14.5, where we indicate that this system is far less understood than Jupiter's. The Uranian satellites are covered in Section 10.14.6, where the regular nature of these satellites in spite of a planetary obliquity of 97° implies that the disk from which these satellites formed was created in the giant impact that led to Uranus' large obliquity – unless the Uranian obliquity resulted from adiabatic processes that allowed the equatorial orbits of the satellites to be preserved. A disk that resulted from a giant impact would make the Uranian satellites the simplest of all giant planet satellite systems to treat, since satellite accretion occurs in a suddenly created disk, and one must worry only about the necessary viscous spreading and cooling of the disk as the satellites form. Alas, no comprehensive theory has yet been constructed. Neptune's satellite system, discussed in Section 10.14.7, is perhaps better understood than most. A plausible series of events centered around the essentially intact capture of the large retrograde satellite Triton accounts well for all of the observational properties of this system. The Pluto–Charon system, like the Earth–Moon system, is characterized by a large specific angular momentum, which likely resulted from an oblique giant impact. In Section 10.14.8 we show how the newly discovered small satellites orbiting in the same plane as Charon could have been created by the same impact that created Charon, where co-formation seems to be dictated by the coplanar orbits.

Minor satellites include the irregular satellites with large semimajor axes, eccentricities, and

inclinations that orbit all of the major planets, and the small close satellites often associated with rings of small particles. The former are distinguished by a capture origin (Section 10.14.9) and the latter by having been broken up and reassembled, probably several times, by collisions among themselves, with comets, or with planetesimals passing through the planets' Hill spheres. The disintegrations of these close satellites are associated with supplying the small particles which make up the observed rings. Many asteroids and Kuiper belt objects (KBO) also have satellites, where the widely separated binaries probably result from capture (Goldreich *et al.*, 2002) and those in close orbits most probably from collisions (Canup, 2005). We will not discuss the small-body satellites beyond this assertion. Rotational fission, precipitation fission, and disintegrative capture among those failed processes offered for the formation of the Moon, also fail to find any likely applications to the other solar system satellites.

10.14.2 The Earth–Moon System

The Moon's mass is 1.2% of the Earth's mass, which is the largest fractional mass of all the satellites in the solar system with the exception of that in the Pluto–Charon system (if we ignore asteroid and Kuiper belt satellites). In addition, the angular momentum of the system of $L_{EM} = 3.5 \times 10^{41} \text{ g cm}^2 \text{ s}^{-1}$ would lead to a 4 h rotation period of the Earth if the Moon were incorporated within. This large specific angular momentum is a key constraint on the origin. The expansion of the lunar orbit by tidal interaction with the Earth (Goldreich, 1966; Touma and Wisdom, 1994) implies formation close to Roche radius $R_R = 2.9R_E$, where R_E is the Earth's radius. Accretion of the Moon inside this distance would be prevented by tidal forces. At a density of 3.34 g cm^{-3} , the Moon is depleted of iron relative to the Earth, and samples of lunar material are depleted in volatiles with vaporization temperatures $< 1300 \text{ K}$ (e.g., Jones and Palme, 2000). A thick layer of anorthositic crust whose constituents have a positive Eu anomaly is complemented with a Eu deficiency in the mare basalts. The thickness of the anorthositic crust coupled with the complementary Eu anomalies is consistent with the crystallization of the lighter anorthositic material in a few hundred kilometer deep magma ocean with the crystals floating to the top. The floating crystals are enriched in Eu leaving the magma, from which the mare basalts are later

derived, deficient in this element (Wood *et al.*, 1970; Wood, 1972; Taylor and Jakes, 1974). But limits on lunar contraction ($< 1 \text{ km}$ in radius) during cooling from the lack of stress features on its surface has been interpreted that the interior of the Moon was initially cold beneath the magma ocean (Solomon, 1986). However, see below for implications of an initially hot moon with reference to a conclusion that contraction might be accommodated without the expected surface scarps. With the exception of the volatiles, similarities in the lunar material with that of the Earth's mantle imply that both the Earth and Moon were formed from material originating at similar distances from the Sun in the protoplanetary disk. The angular momentum and composition constraints are consistent with the formation of the Moon by the impact of a differentiated Mars-sized body with a differentiated Earth (Canup, 2004a).

The formation of the Moon by giant impact was first published by Hartman and Davis (1975) with a lunar-sized impactor that could account for the iron and volatile deficits, but Cameron and Ward (1976) realized that the grazing collision of a Mars-sized body could also account for the large angular momentum of the Earth–Moon system. Models of terrestrial planet accretion (e.g., Wetherill, 1985, 1992) showed that the Earth was likely to have undergone such Mars-sized impacts late in its accretion history, and it is this scenario that has become the paradigm for the formation of the Moon. (For a delightful account of the development of the idea of the giant impact origin of the Moon by the four authors, see MacKenzie (2003); reviewed by Burns (2003).)

The giant impact would have reset the hafnium–tungsten chronometer dating the separation of W from Hf in the Earth's mantle and sending the former to the iron core. Radioactive ^{182}Hf decays to ^{182}W with a half-life $\tau_{1/2} = 9 \text{ My}$. Both elements are refractory, so the planets should have the same overall abundances as the solar system as a whole. But Hf is lithophile (silicate loving) and W is siderophile (iron loving), so during core formation, all the isotopes of W including the radiogenic ^{182}W would be ideally removed from the mantle and sent to the core, while Hf remains in the mantle. So the Hf/W ratio in the mantle would be larger than that in the solar system as a whole, where the latter is inferred from chondritic meteorites. If a planet's core formed with a timescale $< 5\tau_{1/2}$, the mantle would contain excess ^{182}W relative to the other isotopes of W, whereas if the core formed later than this, Hf would have been

extinct and the isotopes of W would be depleted equally leaving the mantle with a chondritic W isotope distribution. The giant Moon-forming impact is the last major event in Earth's accretion (Canup and Asphaug, 2001) so the date of this event should provide the age of both the Earth and Moon. Because the Hf¹⁸²–W¹⁸² ages of the Earth and Moon are model sensitive (especially on the assumed degree of metal-silicate equilibrium during Earth's core formation), ages range from ~ 30 to >100 My after the formation of the solar system (Halliday, 2004; Kleine *et al.*, 2004; Jacobsen, 2005). In contrast, the age of the crystallization of the lunar magma ocean was determined by limiting the Hf¹⁸²–W¹⁸² analysis to metals in returned lunar samples. This choice avoided augmentation of the W¹⁸² from the capture of cosmic ray generated neutrons by Ta¹⁸¹, since Ta is excluded from the metals. The excess W¹⁸² in the metals could thereby be attributed solely to production from Hf¹⁸² decay. A much more narrow range of lunar age deduced from the crystallization age of the lunar magma ocean is determined to be 30–50 My after the formation of the solar system, where the latter is determined by the time of formation of the calcium–aluminum inclusions (CAIs) in some meteorites (Kleine *et al.*, 2005). This Hf¹⁸²–W¹⁸² age is consistent with the formation of the Moon by a Mars-sized giant impact near the end of the Earth's accretion.

The angular momentum relative to the center of mass of the target mass M_t and impactor mass $M_i = \gamma M_T$, with $M_T = M_t + M_i$, the total mass, is (Canup, 2004a)

$$\begin{aligned} L &= b M_T^{5/3} f(\gamma) \sqrt{\frac{2G}{(4\pi\rho_p/3)^{1/3}}} \left(\frac{v_i}{v_{\text{esc}}} \right) \\ &\approx 1.3 L_{\text{EM}} b \left(\frac{M_T}{M_E} \right)^{5/3} \left(\frac{\gamma}{0.1} \right) \left(\frac{v_i}{v_{\text{esc}}} \right) \end{aligned} \quad [1]$$

where $b = \sin \xi$ is the impact parameter divided by $R_t + R_i$, the sum of the target and impactor radii, ξ is the angle between the line connecting the centers of M_t and M_i and the impact trajectory at the time of first contact, such that grazing incidence has $\xi = 90^\circ$ or $b = 1$, G is the gravitational constant, ρ_p is the common density of M_t and M_i , v_i is the velocity of the impactor, $v_{\text{esc}} = \sqrt{2GM_T/(R_t + R_i)}$ is the mutual escape velocity, $f(\gamma) = \gamma(1 - \gamma) \sqrt{\gamma^{1/3} + (1 - \gamma)^{1/3}}$ and M_E is the Earth's mass. The complication of eqn [1] comes from introducing v_{esc} into the equation for angular momentum. For this

simplest case where a single impact provides all of the Earth–Moon system angular momentum, a multi-dimensional parameter space involving b , M_T , γ , v_i must be explored to find those values that yield an appropriately massive, iron depleted Moon.

Smooth particle hydrodynamics (SPH) is the method most used in numerical exploration of the parameter space. The conclusions from these calculations have changed as the spatial resolution has increased and the algorithm and equation of state improved (i.e., from Benz, *et al.*, 1986, 1987 to Canup and Asphaug, 2001; Canup, 2004b). The latest calculations by Canup (2004b) show that for a late impact when the Earth is mostly formed, those impacts that yield more than a lunar mass of iron-poor material outside R_R typically have $b \sim 0.7$, a low impact velocity $v_i/v_{\text{esc}} < 1.1$, mass ratio $0.1 < \gamma < 0.15$ (Mars-sized impactor) for $M_t \sim M_E$, and $L \sim L_{\text{EM}}$ (Canup, 2004a). The material is launched into orbit more by gravitational torques from the distribution of ejected material than by pressure gradients. Predicted temperatures of orbiting material are typically 3000–4000 K with 10–30% vapor by mass. The mean radius of the disk is just outside R_R . Maximum total disk masses range from about 1.5 to 2 lunar masses for $L_i < 1.5L_{\text{EM}}$, which implies accretion of the material into the Moon was fairly efficient.

The SPH calculations are relatively secure during the earlier phases of the impact where the evolution is gravity controlled, and they have the advantage of tracing the ultimate fate of iron versus silicate particles. However, there are still too few SPH particles left in the disk to represent its properties and evolution with confidence, and excess numerical viscosity causes a too rapid spreading of the disk (Canup, 2004a). This means that current SPH calculations are limited to less than about 100 h, the timescale for the SPH viscous spreading of the disk (Canup, 2004a). The first 3-D grid-based, high-resolution hydrodynamic calculations for a Moon-forming giant impact using two extreme equations of state indicate that a completely volatile disk develops shocks that cause the disk material to fall into the Earth, whereas a disk composed mostly of liquid–solid lunatesimals evolves to a steady-state mass outside R_R capable of forming the Moon (Wada *et al.*, 2006). The successful SPH simulations by Canup are consistent with this conclusion, since the resulting disk was only 10–30% volatile, and the SPH calculations are ended before anomalous viscous spreading can become important.

Although the overall compositional and dynamical properties of the current Earth–Moon system seem to be well accounted for by the giant impact origin, a transition from the distribution of material following the impact to the accreted Moon a little outside the Roche radius R_R that is also consistent with current observations has been elusive. The high temperatures of the orbiting material cited above implies a disk that is a mixture of melt and vapor, whose cooling time is longer than the time to completely accrete the Moon (Pritchard and Stevenson, 2000). Even if accretion could somehow be delayed until the disk cooled, N-body calculations (Ida *et al.*, 1997; Kokubo *et al.*, 2000) and analytical estimates (Pritchard and Stevenson, 2000) show such rapid accretion of the cold debris that the initial temperature is again elevated above the melting point. But the lack of thrust faults on the Moon and other indications of the large contraction necessary if the Moon did form hot implies instead that the Moon was initially cold (<1000 K) (Solomon, 1986). On the other hand, uncertainties in the properties of the lunar crust, and in the thermal history seem to allow ranges of parameters and processes that could accommodate an initially hot Moon in spite of the lack of observable indications of a significant radius change (Pritchard and Stevenson, 2000). The most likely consequence of the accretion of the Moon from a disk of material generated by a giant impact is a hot, probably molten Moon orbiting just outside R_R in the Earth’s equatorial plane.

Still, there may be alternatives to this conclusion within the paradigm of a giant impact origin of the Moon. Intact clumps containing a fair fraction of a lunar mass have been observed over all resolutions and all sets of parameter values in the SPH calculations, but most are sheared apart when they pass inside the Roche radius. An example of such an intact clump containing 60% of a lunar mass is shown in Canup (2004b). As these surviving clumps appear to be sheared from the impactor during the off-center collisions, their temperatures may be lower than those particles participating more directly in the collisions. A caveat for this alternative outcome to the giant impact is that SPH simulations are known for spurious clumping (e.g., Imaeda and Inutsuka, 2002).

Before we leave the Earth–Moon system, we must account for the current geometry under the assumption that the semimajor axis of the Moon’s orbit was increased from the initial to the current value by tidal torques. By integrating backwards in time from the

current state, Goldreich (1966) found that Moon’s orbit had an inclination to the Earth’s equator of about 10° when the lunar semimajor axis was approximately $10R_E$. This result was confirmed by Touma and Wisdom (1994). The 10° inclination is not consistent with the Moon having formed in the Earth’s equator plane from the debris generated by a giant impact. Moving the Moon from an equatorial orbit to one inclined by 10° is called the ‘inclination problem’.

There are now four proposals for accomplishing this bit of dynamics. In the giant impact scenario, if the Earth had a significant spin before the impact, and a substantial fraction of the Moon remained in a coherent clump after the impact, the latter’s orbit could have been inclined to the equatorial plane and subsequent accretion of the remaining lunar mass could leave it inclined (Canup, 2004a). Numerical simulations including a large initial rotational angular velocity of the Earth prior to impact have yet to be done. Alternatively, an additional impact on the Earth after the Moon-forming event could have altered the alignment of the Earth’s spin, but this would have to have occurred while the Moon was still close to the Earth (Ward, 2002). Such a second impact so soon after the first is unlikely, however. The remaining two hypotheses increase the initial Moon’s orbital inclination by resonant interactions – one with orbital resonances between the Moon’s motion and the Sun (Touma and Wisdom, 1998), and the other with resonances between the Moon’s motion and a remnant disk inside R_R (Ward and Canup, 2000).

It is perhaps more satisfying to find a way of accounting for the 10° orbital inclination when the Moon was close to the Earth that does not invoke additional events beyond the single giant impact that resulted in the Moon, especially since such impacts are rare events. In an integration that produces the current configuration when the Moon has reached its current distance from the Earth, Touma and Wisdom (1998) start the Moon in the equatorial plane of the Earth at a separation of $3.5R_E$ with an eccentricity of 0.01. The initial obliquity of the Earth is 10° and the initial Earth rotation period is 5 h. There is no circumplanetary disk present. Realistic rates of tidal evolution are used in the symplectic integrations that include the entire chaotic solar system, and dissipation in the Moon is included at a variety of dissipation rates. A tidal model was used employing the constant time lag as discussed in Appendix 2 but with the Mignard (1981a) formulation. The first

strong resonance is encountered when the Moon is at about $4.6R_E$, where the period of the periape motion of the lunar orbit relative to an inertial reference is near 1 year. This resonance is called the evection resonance because the same term in the disturbing function gives rise to the 1.3° amplitude, 31.8 day periodic variation in the Moon's mean longitude called the evection. Capture into the evection resonance is certain if the eccentricity is below about 0.07 as the resonance is approached and if the rate of tidal evolution is sufficiently slow. With the assumed parameters, capture occurs and the eccentricity grows rapidly to large values, where the maximum value reached before the system escapes the resonance is determined by the value of

$$D = \frac{k_M}{k_E} \frac{\Delta t_M}{\Delta t_E} \left(\frac{m_E}{m_M} \right)^2 \left(\frac{R_M}{R_E} \right)^5 \quad [2]$$

where k , Δt , m , R refer to Love number, the constant tidal time lag, mass, and radius with subscripts referring to Earth and Moon, respectively, and with D being a measure of the relative rates of energy dissipation in the Earth and Moon. The current value of D from the lunar laser ranging experiment is about 1.1 (Dickey *et al.*, 1994). For $D=0$ (no dissipation in the Moon) the maximum eccentricity is about 0.5 before escape and the eccentricity continues to climb after escape from the resonance because of tides raised on Earth and no dissipation in the Moon (Goldreich, 1963). For $D=10$ (high dissipation in the Moon) the maximum eccentricity is only about 0.15. For $1 \leq D \leq 10$, the energy dissipated in the Moon in about 8000 years is in the range $\sim 2 \times 10^{35}$ to 1.5×10^{36} ergs, which could lead to substantial melting (Touma and Wisdom, 1998).

After escape from the evection resonance, the continued expansion of the orbit further decreases the prograde motion of the orbit periape and twice the time derivative of the evection resonance variable plus the retrograde motion of the lunar orbit node approaches zero near $6R_E$. The term in the Hamiltonian corresponding to this resonance has $e_M^2 i_M$ in the coefficient, but this resonance affects the inclination more than the eccentricity. Touma and Wisdom name this resonance the eviction – changing the e in evection to i to emphasize the inclination and noting that this resonance ‘evicts’ the Moon from an equatorial orbit. If D is not too large, the eviction resonance is approached with high

eccentricity in the wrong direction for capture. Passage through the resonance leaves the eccentricity large and excites an inclination of $2\text{--}3^\circ$. If D were now to increase drastically – perhaps because the continued high eccentricity has partially melted the interior – the dissipation in the Moon becomes so high that the semimajor axis is decreased as the eccentricity is reduced. This takes the system through the eviction resonance from the other direction where capture and subsequent evolution forces the inclination to values between 9° and 13° . Escape from the resonance is effected because of the continued decrease in e_M , but the remnant inclination is preserved. Subsequent evolution brings the Moon to the current configuration.

Because a remnant disk would contribute to the perigee precession, it must be reduced to less than $0.1M_L$ to bring the evection resonance inside $10R_E$ (Canup, 2004a). A more important constraint on the Touma–Wisdom model is that the Moon begin its tidal evolution cold, which is contrary to our conclusion above that the Moon most likely started hot (e.g., Pritchard and Stevenson, 2000). However, enough uncertainties remain about the transition from the giant impact to an accreted Moon outside R_R that we do not dismiss out of hand this first analysis that allowed the Moon to evolve from an equatorial orbit to its current configuration from the effects of tidal dissipation alone. It is interesting, however, that the recent analysis attributing the observed moment of inertia differences of the Moon to be the frozen remnant of a hydrostatic shape when the Moon was near $25R_E$ distance from the Earth in either synchronous rotation or in a 3:2 spin–orbit resonance requires an initially hot (near solidus or partially molten) Moon, at least in the outer layers (Garrick-Bethell *et al.*, 2006). If this theory prevails, it may preclude the above dance through resonances to account for the necessary increase in the orbital inclination to a value near 10° . But the inference of reaching a highly eccentric orbit in spite of the higher dissipation in a soft Moon must be reconciled with the limited eccentricity determined for such a Moon that required it be initially cold to allow the resonance dance.

The scheme involving disk interactions to change the Moon's inclination (Ward and Canup, 2000) depends on the coexistence of the Moon with a 0.5–1 lunar mass Roche-interior disk for 10–100 years. Waves in the disk are generated at locations where the ratio of the local mean motion to that of the Moon involves two small integers. Torques that

result from the gravitational interaction of the Moon with the distribution of mass in the wave structures can affect the lunar orbit. Such mean motion resonances from planar interactions are called Lindblad resonances, and they generate spiral density waves. Spiral bending waves are associated with the inclination of the Moon's orbit, where material in the disk is lifted out of the plane (e.g., Shu, 1984). When the Moon is just outside R_R , there are many mean motion resonances in a disk that extends nearly to R_R . Torques from the waves generated in the disk force the Moon to larger semimajor axes while the back reaction reduces the angular momentum contained in the disk. The resonances thereby move out through the disk and eventually leave its outer edge until the last strong one to leave the disk is the 3:1 inner vertical resonance (IVR) when the Moon's semimajor axis is a little over twice the disk radius. The effect of the torque $T_{\text{IVR}} \propto \sigma_g \sin^2 I$ from the so-called bending waves, where I is the orbital inclination for the Moon and σ_g is the surface mass density of the gas in the disk, in addition to contributing to growth of the semimajor axis a of the Moon's orbit, is to increase the inclination at the rate (Ward and Canup, 2000),

$$\frac{dI}{dt} = \frac{T_{\text{IVR}}}{M_M a^2 \Omega_K \sin I} \left(\frac{3}{2} \cos I - 1 \right) \quad [3]$$

where M_M is the lunar mass and Ω_K is the Kepler mean motion. With a disk mass of $0.75 M_L$ and an initial $I = 1^\circ$, which initial condition is motivated by lunar formation models (Canup and Esposito, 1996; Ida *et al.*, 1997), inclinations reached 12.3° and 14.5° as the Moon nears $6R_E$ for initial disk spreading times of 37 and 50 years, respectively (Canup, 2004a). The growth of I stops if either the resonance moves outside the disk, or if the disk is depleted. The disk interaction mechanism requires control of the evolution by the 3:1 vertical resonance. Other resonances, if still present, can lead to inclination damping (e.g., Borderies *et al.*, 1984; Ward and Hahn, 1994). However, the Lindblad resonances invoked by Borderies *et al.* to damp the inclination are out of the disk when the 3:1 vertical resonance is near the disk edge, and the co-orbiting vertical resonances used by Ward and Hahn are at the position of the Moon, which is outside of the disk. Once the inclination is at 10° near the Earth, tidal evolution can carry the Moon to the current distance with the orbit inclined by about the observed 5° relative to the ecliptic.

The massive disk makes the Ward–Canup and the Touma–Wisdom schemes mutually exclusive because the higher precession rate of the periapse makes the ejection and eviction resonances further away than the distances where Ward–Canup evolution takes place. If the Moon is partially molten in the Ward–Canup scenario, the high dissipation limits the eccentricity growth in a subsequent passage through the ejection resonance as the disk is dissipated and the lunar orbit continues to expand, even if trapped for a time (Touma and Wisdom, 1998). The passage through the eviction resonance is in the wrong direction for capture so the inclination will not change significantly after the Ward–Canup mechanism has brought it above 10° .

10.14.3 Mars System

Mars' inner satellite Phobos is located at a distance of 2.77 Mars radii (R_M) from the center of Mars, which is well inside the co-rotation radius of $\sim 5.9R_M$, and Deimos is just outside the co-rotation radius. The tides raised on Mars thus cause Phobos to be spiraling toward Mars and Deimos to be spiraling away. In fact Phobos is inside the Roche radius for a density of 1.9 g cm^{-3} and would be torn apart by tidal forces if it were a fluid. It needs a shear strength of only $10^5 \text{ dynes cm}^{-2}$ to resist disruption (Yoder, 1982) – a loose rubble pile would survive (Soter and Harris, 1977; Dobrovolskis, 1982).

Determinations of the dissipation function Q (Appendix 2) of Mars are $O(100)$ (Shor, 1975; Sinclair, 1978; Duxbury and Callahan, 1981; Bills *et al.*, 2005) from observations of the secular acceleration of Phobos' orbital mean motion. If we choose a constant value of $Q_M = 100$ with Love number $k_M = 0.16$ (Bills *et al.*, 2005), Deimos' orbit could have expanded by less than 200 km in 4.6×10^9 years. Phobos' initial semimajor axis would have been near $5.6R_M$ under the same assumptions. The rotation period of Mars has been essentially unaffected by the exchange of angular momentum with the satellites and would have been increased by only about 10 min due to solar tides. Deimos has essentially its initial orbit, and Phobos having started inside the co-rotation radius is consistent with the measured current value of $Q_M \approx 100$.

Both Phobos and Deimos are synchronously rotating, Deimos would have reached this state in less than 10^8 years from an unlikely small initial spin period of 4 h, where a rigidity of $5 \times 10^{11} \text{ dynes cm}^{-2}$

and $Q=100$ were assumed. Under the same assumptions Phobos would have reached this state in less than 10^5 years at its current separation from Mars and in less than 10^7 years at its likely initial separation near the co-rotation radius (see Peale (1977) for a detailed discussion of the rotation histories of all of the satellites known at that time).

The two satellites of Mars are in nearly circular equatorial orbits and this argues for their accretion *in situ* from a debris disk in the equatorial plane left over from the planet's formation – like the regular satellites of the major planets. However, Phobos' mean density was estimated to be $2.0 \pm 0.5 \text{ g cm}^{-3}$ from Viking orbiter data (Christensen *et al.*, 1977; Tolson *et al.*, 1978), and that of Deimos $2.0 \pm 0.7 \text{ g cm}^{-3}$ (Duxbury and Verkra, 1978). Another estimate for the density of Phobos comes from $GM_p = 7.22 \pm 0.05 \times 10^{-4} \text{ km}^3 \text{ s}^{-2}$ determined from the 5 day rendezvous by Soviet spacecraft Phobos 2 (Kolyuka *et al.*, 1990) coupled with the volume of $5748 \pm 190 \text{ km}^3$ from Thomas (1993) to yield a density of $1.88 \pm 0.07 \text{ g cm}^{-3}$. Some more recent estimates of the density from cumulative distant encounters by Mariner 9 and Viking 1 and 2 spacecraft are even lower at $1.53 \pm 0.10 \text{ g cm}^{-3}$ for Phobos and $1.34 \pm 0.83 \text{ g cm}^{-3}$ for Deimos (Smith *et al.*, 1995). Yuan *et al.* (2001) find $GM_p = 7.138 \pm 0.019 \times 10^{-4} \text{ km}^3 \text{ s}^{-2}$ for Phobos and $GM_D = 1.497 \pm 0.105 \times 10^{-4} \text{ km}^3 \text{ s}^{-2}$ for Deimos, which yield densities of 1.86 ± 0.07 and $2.21 \pm 0.45 \text{ g cm}^{-3}$ for Phobos and Deimos, respectively, where $V_p = 5748 \pm 190 \text{ km}^3$ and $V_D = 1017 \pm 130 \text{ km}^3$ (Thomas, 1993) are used. All the estimates for the Phobos and Deimos densities are $\sim 2 \text{ g cm}^{-3}$ except the Smith *et al.* value, which is somewhat lower. The difference from Mars' mean density of 3.9 g cm^{-3} lends support to the suggestion that the satellites are composed of material which did not originate in the vicinity of Mars, although porosity is another means of reducing the density. However, albedos near 5% and reflection spectra are consistent with carbonaceous chondritic material (Pang *et al.*, 1978; Pollack *et al.*, 1978). Intact capture was considered (Pollack and Burns, 1977; Mignard 1981b), but Szeto (1983) showed several seemingly insurmountable inconsistencies with the capture hypothesis independent of the impossibility of relaxing the captured satellites into the circular equatorial orbits where they are found. In spite of the low density and spectral indications of carbonaceous chondrite material, the coplanar, equatorial orbits indicate that these satellites must have formed from a dissipative disk

of debris orbiting the planet. If that debris were indeed carbonaceous chondritic, which is not at all certain, one possible way it could have gotten into orbit about Mars would be the shattering of such a planetesimal that was formed in the asteroid belt region of the nebula when it collided with a denser object already in orbit about Mars. The condition here would be that the pieces would have to be sufficiently small and of sufficient number to make a dissipative disk. Samples of both Phobos and Deimos would tell us if such a contrived origin were necessary. Any scheme to capture these satellites intact and bring them into their current orbits cannot survive close inspection of the assumptions involved.

An interesting alternative to placing carbonaceous material into a dissipative disk around Mars by collision with an existing small satellite is a collision with Mars itself by a much larger body, as much as 1800 km in diameter (Craddock, 1994). Like the giant impact creating the Moon and accounting for the large angular momentum of the Earth–Moon system, this impact could account for the relatively high rotation rate of Mars, which is difficult to achieve without a large impact (Dones and Tremaine, 1993), while distributing debris in the equatorial plane and perhaps leaving a record of the impact in the 7700 km Borealis basin. If the impactor were carbonaceous, much of the orbiting material would be this type. Many Phobos-like objects may have impacted Mars leaving a distribution of craters with asymmetric or elongate ejecta (Schultz and Lutz-Garhan, 1982), and we are seeing only the surviving members of the swarm.

The hypothesis that Phobos and Deimos are remnants of material in a dissipative disk, whatever the disk's origin, is not compromised by the inferred orbital history of the satellites. The initial equatorial orbits remain equatorial in spite of the chaotic, large amplitude variations in the obliquity of Mars (Ward, 1979; Laskar and Robutel, 1993; Touma and Wisdom, 1993) and despite the precession of the spin axis of Mars (Goldreich, 1965). The solid angle described by the orbit normal as the satellite orbit precesses due to Mars' oblateness is an adiabatic invariant (Goldreich, 1965), as the precession rates for the Martian satellites (periods of 2.3 and 57 years for Phobos and Deimos, respectively; Peale, 1977) are fast compared with rates of change of Mars' spin axis direction relative to inertial space (timescales $O(10^5)$ years); Touma and Wisdom, 1993; Folkner *et al.*, 1997).

If we insist that both satellites started with nearly circular orbits, how then can we explain the current eccentricity of Phobos' orbit $e_p = 0.0151$? If the orbital motion is integrated backward in time, this eccentricity grows to large values and collisions with Deimos would have been likely (Yoder, 1982; Szeto, 1983), even if there were no tidal dissipation in Phobos. Significant dissipation in Phobos reduces the timescale for a crossing orbit with Deimos to less than 10^9 years in the past (Yoder, 1982). The current eccentricity cannot therefore be a remnant eccentricity from tidal decay beginning 4.6×10^9 years ago. Yoder (1982) has identified three commensurabilities (defined when two characteristic periods in the description of the motion are in the ratio of small whole numbers) that Phobos has passed through within the past 10^9 years that provide likely gravitational excitations of Phobos' eccentricity during its inward spiral. The commensurabilities are encountered at semimajor axis $a = 3.8, 3.2$, and $2.9R_M$, where the earliest resonance was encountered only 5×10^8 years ago. The first and third are 2:1 and 3:1 commensurabilities between the rotation of Mars and the orbital mean motion, where the resonant interaction is with Mars' axial asymmetry. At $3.2R_M$, the 2:1 commensurability is between the apparent mean motion of the Sun and the periape motion of Phobos's orbit, where the latter's secular motion is due to the oblate figure of Mars. This resonance is like the evection resonance for the Moon. There is also a 3:2 spin-orbit resonance excitation of the eccentricity when $a = 4.6R_M$, but this excitation happened so long ago that there would be no contribution to the current eccentricity. The eccentricity would have decayed after each excitation and plausibly arrives at the current eccentricity after the series of kicks and subsequent decays (Yoder, 1982). Orbital inclination can also be excited, and even though the resonance interaction is not as strong as it is for the eccentricities, the excited inclinations do not decay. Still, the current inclinations of the orbits are consistent with the resonance passages (Yoder, 1982).

There is a condition on the dissipation in Phobos for this scenario to work. Yoder (1982) has calculated the dissipation in the satellite accounting for both the tidal dissipation due to eccentric orbit as discussed earlier and that due to the forced libration of the very asymmetric satellite. This libration has an amplitude of 3.9° (Duxbury and Callahan, 1981; Yoder, 1982) and causes twice the tidal dissipation in Phobos that would occur if Phobos were nearly axially symmetric in the same eccentric orbit (Yoder, 1982). Both the dissipation in Phobos and that in Mars due to

tides raised by Phobos damp the eccentricity (see eqn [58] in Appendix 2). There cannot be too much damping since the series of eccentricity excitations or the current eccentricity would be less than that observed. Since the dissipation in Mars can be presumed known from the measurement of Q_M , and the magnitude of the probable excitations can be reasonably estimated from the resonance passage analysis, the current value of e_p limits the contribution by Phobos. Yoder finds that $\mu_p Q_p > 3-6 \times 10^{12}$ dynes cm $^{-2}$ or, if $Q_p \sim 100$, $\mu_p \gtrsim 10^{10}$ dynes cm $^{-2}$, which is about that of ice. The properties of Phobos are not sufficiently well known to be sure that the rigidity could satisfy this constraint, but this rigidity is not unreasonable.

During Phobos' spiral toward Mars, it is likely that it passed through the 2:1 orbital mean motion commensurability with Deimos. Such a passage would excite an eccentricity of about 0.002 in Deimos orbit if the eccentricity of Deimos were much smaller than this before resonance passage (Yoder, 1982). The time of this commensurability is known if the current dissipative properties of Mars have not changed substantially since the resonance encounter. This places a lower bound on the dissipation in Deimos if the current eccentricity (0.0002) is the tidal remnant from an initial value of 0.002 excited by the resonance passage. Yoder (1982) finds $\mu_D Q_D (1 - \alpha_D)^2 / \alpha_D^2 \gtrsim 10^{10}$ dynes cm $^{-2}$, where $\alpha_D = 3(B - A)/C$ with $A < B < C$ being the principal moments of inertia of Deimos. This limit may be unreasonably low, but the dissipation in Deimos may be increased if the forced libration is nearly resonant with the free libration. The enhanced amplitude of libration would lead to higher dissipation and relax the constraint on $\mu_D Q_D$. The free libration period could be better constrained by an estimate of α_D from a more accurate determination of Deimos' shape along with an accurate measure of its physical libration amplitude.

In any case, Yoder's hypothesis that the satellite orbits have always been regular and current properties of the system then attributed to the effects of resonance passages by Phobos is well supported. This is consistent with our presumed origin from a dissipative disk of small particles.

10.14.4 Jupiter System

The Galilean satellites of the Jupiter system and those small satellites inside Io's orbit are nearly coplanar with Jupiter's equator implying their

formation in a dissipative disk (Appendix 1). Additional constraints on the formation of the Galilean satellites of Jupiter have been established by theoretical analysis and the Galileo spacecraft observations. The mean densities of Io, Europa, Ganymede, and Callisto are, respectively, 3.53, 3.01, 1.94, and 1.83 g cm^{-3} . Ganymede and Callisto are about 40% ice and 60% rock by mass (Sohl *et al.*, 2002; Chapter 10.15). Europa has a layer of water and ice 80–170 km thick (Anderson *et al.*, 1998b). CO₂ has been detected on the surfaces of Ganymede and Callisto (McCord *et al.*, 1998; Hibbitts *et al.*, 2000, 2002, 2003) and in the atmosphere of Callisto (Carlson, 1999). The discovery of an intrinsic dipole magnetic field on Ganymede and the low amplitude of higher order multipoles in the field imply generation of the field deep within the satellite probably by dynamo action in a molten metallic core (Kivelson *et al.*, 1997, 1999, 2002; Chapters 10.07, 10.09, and 10.13). Induced fields in Europa and Callisto and possibly Ganymede from the variation in the external field due to Jupiter's rotation are consistent with induced fields expected from a conducting liquid water layer at rather shallow depths (Kivelson *et al.*, 2002). It has been speculated that NH₃ is sufficiently abundant on Ganymede and Callisto to lower the solidus temperature of an NH₃–H₂O mixture sufficiently so that such a liquid layer could be preserved over the age of the solar system (Mousis *et al.*, 2002b; Spohn and Schubert, 2003). However, there has been no NH₃ nor N₂ detected on any of the Galilean satellites, but McKinnon (2006) argues that halides in solution are sufficient to depress the solidus temperature sufficiently to preserve the liquid layers.

The large amount of relatively volatile material in Ganymede and Callisto implies that the disk temperature remained sufficiently cool during the accretion process that icy particles could persist. Callisto may not be fully differentiated (Anderson *et al.*, 1998a, 2001; Nagel *et al.*, 2004), which requires an accretion timescale $\sim 10^5$ – 10^6 years to keep temperatures below 273 K (Stevenson *et al.*, 1986; Stevenson, 2001a). The change in composition from rocky to icy from Io to Callisto is consistent with an expected radial temperature gradient in the disk at the time of satellite accretion. The effectiveness of the possible removal of an initial major mass fraction of volatiles from Io and Europa from tidal heating (Appendix 2) has not been ascertained.

The formation of the regular Jovian satellites must necessarily occur late in Jupiter's accretion to allow the hot inflated protoplanet to cool and contract to a

radius less than that of the closer satellites. Minimum mass disk models (called minimum mass sub-nebula models (MMSN)) for the formation of the satellites require sufficient mass of condensable material to form the satellites augmented with gas to reach solar composition in the initial disk. No mass is added to the disk during the formation (Lunine and Stevenson, 1982; Mosqueira and Estrada, 2003a, 2003b). But because of the short accretion times in the MMSN disk, indicated in Appendix 1, the low temperatures implied by the current abundance of ice in the satellites, and the rapid migration of large satellites in a massive disk, models where the Galilean satellites form during the very last stages of Jupiter's accretion when the accretion disk was starved (Stevenson, 2001a, 2001b; Canup and Ward, 2001, 2002, 2006; Alibert *et al.*, 2005a; Mousis and Alibert, 2006) comprise a more likely scenario. In these latter models considerably more condensable mass passes through the disk than is contained in the current satellites.

In spite of Jupiter having opened a gap (Appendix 1) in the solar nebula when its mass exceeded approximately one-third of its current value (e.g., Bryden *et al.*, 1999; Ward and Hahn, 2000), 3-D numerical simulations (e.g., Kley *et al.*, 2001) find continuing accretion onto Jupiter at a rate of 10^{-2} – $10^{-4} M_{\oplus} \text{ yr}^{-1}$, depending on the assumed nebular viscosity, surface mass density, and scale height. Bate *et al.* (2003), in a 3-D hydrodynamic simulation, find an accretion rate close to $10^{-4} M_{\oplus} \text{ yr}^{-1}$ ($3 \times 10^{-2} M_{\oplus} \text{ yr}^{-1}$) as Jupiter approaches its final mass. The latter authors find that accretion only stops when the planet at 5.2 AU reaches about $10 M_{\oplus}$, with an undepleted solar nebula. The accretion of Jupiter therefore must be protracted with the rate of accretion tailing off as the solar nebula is depleted.

The shocks near the L_1 and L_2 Lagrange points, after a gap is formed in the solar nebula, which are found in the 2-D calculations of Bryden *et al.* (1999), Kley (1999), and Lubow *et al.* (1999), are much weakened in the 3-D calculations of Bate *et al.* (2003). In the final stages of accretion during the starved phase of the Jupiter disk history, the circumplanetary disk is more likely to behave like a standard accretion disk rather than being subject to the shock-driven accretion found in the 2-D calculations (Bate *et al.*, 2003). In both 2-D and 3-D calculations, a prograde accretion disk is formed, which is the basis for the formation of a coplanar set of satellites in nearly circular orbits. A striking conclusion from the 3-D calculations of Bate *et al.* (2003) is that after the gap is formed, relatively little material from the mid-plane

of the solar nebula is accreted onto the disk. Most of the material rains down on the disk from above and below the mid-plane. Accounting for constraints from observations, numerical simulations and logical thoughts about the details of planet formation, we shall favor the starved disk models, where the satellites accumulate during the last stages of planetary accretion. But first we describe a currently advocated minimum mass model and a gas-free accumulation of the satellites from planetesimals captured from the heliocentric swarm.

10.14.4.1 Minimum Mass Sub-Nebula (MMSN)

Mosqueira and Estrada (2003a, 2003b) have adopted a MMSN hypothesis in which to form the regular satellites with various schemes for avoiding the short timescales for the important processes in the disk discussed in Appendix 1. The system is closed, since continuing accretion is omitted. A $1/r$ temperature distribution is assumed with $T = 250\text{ K}$ at Ganymede's distance in order to accommodate the stability of icy particles at greater distances. A unique feature of the model is a surface density distribution ($\sigma_g \propto 1/r$) with sufficient solids to make Io, Europa, and Ganymede inside $\sim 20 R_J$ leading to an optically thick, massive inner disk ($\sigma_g \sim 10^5 \text{ g cm}^{-2}$). Mosqueira and Estrada set the value of the gas surface density based on that required for the Galilean satellites to open gaps, leading to $\sigma_g \sim 3 \times 10^4$ to $3 \times 10^5 \text{ g cm}^{-2}$, with the lower end of the range indicating a disk enhanced in solids compared to the standard MMSN disk. But the surface density falls drastically beyond about $20 R_J$ to a value determined by spreading Callisto's mass in the outer disk out to $r \sim 150 R_J$. This results in an optically thin outer disk from which Callisto is accreted slowly ($\sigma_g \sim 1/r$ for $r > 26 R_J$). The $1/r$ temperature dependence in the inner disk goes to a constant temperature in the outer disk of 130 K , like that of the solar nebula. The surface mass distribution is justified by the assumption that most of the material accreted onto the disk near the end of the accretion phase will fall inside a 'centrifugal radius', where the latter is determined from the total specific angular momentum of solar nebula material relative to Jupiter that encounters the Roche lobe of Jupiter, and conserving that angular momentum as the material is transferred to the circumplanetary disk (Lissauer, 1995). The centrifugal radius adopted by Mosqueira and Estrada is $R_H/48 = 15R_J$ for Jupiter. A

controversial assumption in the model is that the sharp gradient in the surface density between the inner and outer disks is maintained throughout the accretion timescale of Callisto, which is $O(10^5 - 10^6 \text{ years})$. The low density and temperature of the outer disk allows a sufficiently slow accretion of Callisto to be consistent with a possible lack of complete differentiation.

The steep surface density gradient between inner and outer disks is maintained by assuming that turbulence has been nearly completely damped leading to $\alpha \sim 10^{-5} - 10^{-6}$ in both inner and outer disks. Such low values of α might not be definitely ruled out, since there is no continuing accretion to stir the disk. Small but decoupled objects are rapidly lost to Jupiter through gas drag. But the accumulation of dust and rubble into larger objects is more rapid than the decay of these objects by gas drag. The small particles are assumed entrained in the gas and encounter larger particles that are at Kepler velocities. The timescale is $M/\dot{M} = 4\rho_p R/(3\rho_s \Delta v)$, where Δv is the difference between the Kepler and gas velocities. Sweepup time is smaller than gas drag time for $r < 38R_J$ independent of particle size or distance where $T \propto 1/r$, such that satellite formation is favored. Then the timescale for accretion is determined by the rate at which gas drag can bring say 100 m or larger objects to the protosatellite. Accretion times for Ganymede under these assumptions are $O(10^3 - 10^4 \text{ years})$. No constraints are placed on the accretion timescales of Io and Europa because of their low volatile content.

The short migration timescales indicated in Appendix 1 are frustrated by the assumption that the inner three Galilean satellites open gaps in the disk and change from type I to type II migration. The latter is on the viscous timescale, which exceeds the supposed timescale for elimination of the gas disk by photo evaporation (Shu *et al.*, 1993; Hollenbach *et al.*, 2000), because the turbulent viscosity characterized by $\alpha = O(10^{-6})$ is so small. The steep gradient in the surface mass density between 20 and $26R_J$ causes type I drift to reverse sign in this region so that a satellite would spiral out instead of in. Callisto is hypothesized to have ended up where the type I drift vanishes at the value of the density gradient where torques from the inner Lindblad resonances balance those from the outer ones. The steep gradient in σ_g between the inner and outer disks must not only be maintained through the accretion phase, but also throughout the dissipation of the disk. Alternatively, Callisto could open a gap, shift from type I to type II drift, and thereby end up where it

is because of the low viscosity assumed. But why did it wait until $r=26R_J$ to open the gap?

Many of the subjective criticisms of the Mosqueira–Estrada model, such as the very low values of α assumed from the outset to preserve the steep surface density gradient between the inner and outer disks, are muted by uncertainties about which processes are ongoing and on the values of various parameters therein. However, many aspects of the model are *ad hoc*, such as the $1/r$ temperature distribution normalized to allow ice to condense just beyond Ganymede's orbit, the low viscosity (and maintenance of the steep surface density gradient between inner and outer disks), and the assumption of the existence of a minimum mass, closed disk in the first place. This and all closed disk models for the formation of the satellites around the giant planets are not placed in the context of accretion of the planet. Recent results from increasingly sophisticated numerical simulations of planetary accretion (e.g., Bate *et al.*, 2003), which show the protracted nature and persistence of the accretion process as the solar nebula is dissipated, are ignored. These results of the numerical simulations confirm intuitive conclusions about the accretion histories of the giant planets. As a consequence MMSN models are not likely to represent the actual sequence of events that led to the formation of the regular natural satellites around the giant planets. In spite of these criticisms, we note that the Mosqueira–Estrada papers contain discussions of many if not most of the processes that occur in disks that effect the satellite development. These processes should be studied in the development of a comprehensive model of origin of the regular satellites. Especially intriguing are comparisons of the Jupiter, Saturn, and Uranus systems relative to their respective Hill sphere radii.

10.14.4.2 Planetesimal Capture Model

In a rather detailed and careful analysis, Estrada and Mosquiera (2006) re-examine the hypothesis that collisions of planetesimals within Jupiter's Hill sphere in a gas-free environment led to a buildup of captured orbiting material that settles to an equatorial disk through mutual collisions where they could have accumulated into the satellites (e.g., Safronov *et al.*, 1986). Constraints are placed on numerous parameters and assumptions that could allow such a scenario to produce the observed masses and angular momentum distributions, but the authors note that whether or not the parameter values and assumptions

are likely to obtain is very uncertain. Perhaps the greatest value of this paper is to show how difficult it is to make a gas-free planetesimal capture scenario a viable alternative for satellite formation. Major caveats are the difficulty in producing the composition gradient in such a scenario (unless later Roche lobe interlopers or tidal dissipation preferentially removes volatiles from the inner satellites), keeping a size distribution of planetesimals that favors collisional capture, and maintaining the supply of planetesimals within Jupiter's feeding zone over the 10^5 – 10^6 year timescale of the satellite accumulation. It is not clear that a sufficient gas density in the solar nebula to allow a gas drag replenishment of planetesimals in Jupiter's feeding zone would also not keep a gas accretion disk replenished by streaming across the gap. A problem with planetesimals later colliding with the accreted satellites to preferentially remove volatiles from the inner satellites is attaining a sufficient flux of such planetesimals from the depleted zone around Jupiter.

Another problem that is not addressed is the reduction of the optical depth in the circumplanetary disk as the satellites accumulate into larger objects on short timescales thereby reducing the collisional capture probability of additional mass. In addition, the scenario is not placed in the context of the formation of Jupiter that yields a plausible route to the assumed gas-free initial conditions where Jupiter has no satellites. Presumably, Jupiter has already gone through the last stages of gas accretion through streams of material across a gap in the solar nebula, and we shall see below that it should have acquired a stable satellite system already during that process. With the caveats pointed out by Estrada and Mosqueira and the additional ones we have mentioned here, it seems unlikely that the satellites formed solely from collisionally captured planetesimals in a nearly gas-free environment.

10.14.4.3 Starved Accretion Disk Model

Accumulation of the satellites of Jupiter in a starved accretion disk instead of a MMSN was first advocated by Canup and Ward (2001) and separately by Stevenson (2001a) at the Jupiter Conference in Boulder Colorado in June 2001. The motivation for the starved disk model comes from the necessity of disk temperatures low enough to allow water ice to condense and to allow sufficiently slow accretion of Callisto to allow possibly incomplete differentiation (e.g., McKinnon, 1997; Kuskov and Kronrod, 2005).

Stevenson (2001b) summarized remarks at the conference, whereas Canup and Ward (2002) have published a detailed steady-state model based on their conference paper. Their model was extended by Alibert *et al.* (2005a, 2005b) who account for the time variation of disk properties as accretion wanes. Major features in the Alibert *et al.* approach are that material from the solar nebula is deposited at the outer edge of the disk with a mass flux through the disk that is in quasi-steady state, and that the last part of the satellite formation occurs in a closed disk after accretion has ended. Canup and Ward (2002) have the mass raining down onto the disk consistent with a major finding of the 3-D calculations of Bate *et al.* (2003) showing the vertical deposition. Canup and Ward (2006) have also extended their model to include an exponential decrease in the accretion rate with time, where many of the accreted satellites are lost to Jupiter from migration and only the last set formed near the end of accretion remain. It is the Canup–Ward model (2002, 2006) that we now describe.

For accretion during gas inflow, the steady-state disk with uniform deposition inside a radius r_c treated in 2002 is generalized in 2006 to a time-dependent disk with a flux density onto the disk surface inside radius r_c given by

$$F_{\text{in}} \propto \left(\frac{1}{r}\right)^{\gamma_{\text{in}}} \exp\left(-\frac{t}{\tau_{\text{in}}}\right) \quad [4]$$

where F_{in} is the vertical flux density onto the disk's upper and lower surfaces, r is the radial distance from the center of Jupiter, τ_{in} is the time constant for exponential decay of the influx rate, and γ_{in} is an adjustable constant parameterizing the radial variation in the flux density over the disk inside a radius r_c . The radius r_c is a free parameter constrained by the specific angular momentum of inflowing material. A range of 16–46 planetary radii for Jupiter, Saturn, and Uranus is consistent with Lissauer's (1995) 3-D calculation (with 16 being the value for Jupiter in particular). The disk spreads viscously, and loses mass to Jupiter and at an outer boundary $r_d \approx 150R_J$, where material is stripped from the disk. If the viscous timescale $\tau_v \approx r^2/v$, with $v = \alpha c_s H$ (c_s = mid-plane sound speed, H = gas scale height, Appendix 1), is much smaller than τ_{in} , the gas surface density σ_g maintains a quasi-steady state like that assumed in Canup and Ward (2002). Solids smaller than ~ 1 m are entrained in the gas and delivered with it to the disk from heliocentric orbit.

Once in the disk, the solids accrete rapidly enough with timescale $\tau_{\text{acc}} < \tau_{\text{GD}}$ before they can spiral into Jupiter from gas drag. The accretion timescale for a satellite of mass m_s is $\tau_{\text{acc}} \approx f m_s / (2\pi r \Delta r F_{\text{in}})$, where $f \approx 100$ is the gas to solid ratio for the influx, and Δr is the width of the annulus over which the satellite accumulates material. With $\Delta r/r \approx 2e$, where $e \approx (H/r)(m_s/4\pi r H \sigma_g)^{1/3}$ is the maximum eccentricity obtained by m_s from a balance between eccentricity damping by density waves and excitation by scattering from similarly sized objects (Ward, 1993). Satellite growth continues at a rate that is controlled by the influx rate, while each satellite spirals toward Jupiter from Type I drift (Appendix 1). The rate of Type I drift is proportional to the satellite mass and to the gas surface density in the disk (Appendix 1), so for the values of surface density gradients obtained in the model, the satellite drift toward Jupiter accelerates as the mass grows and as the surface density increases at smaller distances from Jupiter.

The timescales for Type I drift (Appendix 1) are generally shorter than τ_{in} in eqn [4] for the waning influx, so that many satellites are lost to Jupiter as the influx winds down. The satellites reach a critical mass m_{crit} where the timescale τ_{acc} for further growth is comparable to its Type I orbital decay timescale after which further growth is limited before the satellite is lost to Jupiter. The model includes disk heating and cooling, where the heating comes from the gravitational energy lost by the incoming material, viscous dissipation within the disk, and even radiation from a still warm Jupiter with only radiative cooling to keep the disk temperature below the ice point at least in the outer regions. Disk temperature distribution and surface densities are constrained by assumptions about the disk opacities and the viscosity parameter α defined in Appendix 1. The timescale for Type I drift depends on the scale height, which depends on the temperature of the disk, and that temperature distribution depends on the opacity of the solids–gas mixture in the disk, which itself is a function of temperature (Pollack *et al.*, 1994) (Appendix 1). Higher opacities lead to slower required inflow rates for similar temperatures. The importance of the opacity in determining disk structure and the migration of the satellites therein cannot be overemphasized. Unfortunately, the opacity and its dependence on radial distance from Jupiter is highly uncertain, since it depends on the amount, composition, and size distribution of the dust in the nebula. Canup and Ward (2002, 2006) use a constant opacity in their model, and it would be of interest to

determine the effect of other assumptions about the opacity distribution on their conclusions.

For the case where $\gamma_{\text{in}}=0$ in eqn [4] (uniform deposition inside r_c) Canup and Ward (2006) find in their eqn [2]

$$\begin{aligned} \frac{m_{\text{crit}}}{M_J} &\approx 5.4 \left(\frac{\pi}{C_a} \right)^{5/9} \left(\frac{H}{r} \right)^{26/9} \left(\frac{r}{r_c} \right)^{10/9} \left(\frac{\alpha}{f} \right)^{2/3} (\Omega_k \tau_G f)^{1/9} \\ &\approx 5.6 \times 10^{-5} \chi \left(\frac{3.5}{C_a} \right)^{5/9} \left(\frac{H/r}{0.1} \right)^{26/9} \\ &\quad \times \left(\frac{r/r_c}{0.5} \right)^{10/9} \left(\frac{\alpha/f}{3 \times 10^{-5}} \right)^{2/3} \end{aligned} \quad [5]$$

where $\chi = [(1 \text{ week}/\{2\pi/\Omega_k\})(\tau_G/10^7 \text{ yr})]^{1/9}$, $\tau_G = M_J/(dM_J/dt)$ is the timescale for the growth of M_J . Because the ratio m_{crit}/M_J depends so weakly on the inflow rate ($\tau_G^{1/9}$), a similar maximum satellite mass results for a wide range of inflow rates. Next, the dependence of m_{crit}/M_J on α/f for a given inflow rate can be understood from the following. A larger α leads to a lower surface mass density and hence slower migration, which allows the satellite mass to grow more before it is lost, and a smaller f means more solids are delivered also resulting in an increased satellite mass.

In these analytical approximations, each satellite grows to mass $m_{\text{crit}}(r)$ within an annulus of width Δr , after which it is rapidly lost from Type I migration. But in the time to migrate into Jupiter, another m_{crit} accumulates. So the total mass M_T in satellites at any time is given by eqn [4] of Canup and Ward (2006),

$$\begin{aligned} \frac{M_T}{M_J} &= \int_{R_J}^{r_c} \frac{m_{\text{crit}}/M_J}{\Delta r} dr \\ &\approx 2.1 \left(\frac{\pi}{C_a} \right)^{4/9} \left(\frac{H}{r} \right)^{10/9} \left(\frac{\alpha}{f} \right)^{1/3} \frac{1}{(\Omega_k \tau_G f)^{1/9}} \\ &\sim 2.5 \times 10^{-4} \frac{1}{\chi} \left(\frac{3.5}{C_a} \right)^{4/9} \left(\frac{H/r}{0.1} \right)^{10/9} \left(\frac{\alpha/f}{3 \times 10^{-5}} \right)^{1/3} \end{aligned} \quad [6]$$

where it is assumed that H/r and f are approximately constant across the disk and that $R_J \ll r_c$. An important characteristic of eqn [6] is that M_T/M_J is comparable to the observed satellite–planet mass ratios for Jupiter with plausible choices for the parameters, and by replacing M_J by M_S or M_U , the same applies to Saturn and Uranus. Notably, M_T/M_J is insensitive to the inflow rate through χ , lacks a dependence on r_c and depends only weakly on α/f .

Perhaps the most notable result of this model is the prediction that all the major planets should have values of $M_T/M_P \approx 2 \times 10^{-4}$ (M_P is the planet mass),

which is what is observed. The value of M_T/M_P is determined primarily by the value of α/f , and the weak dependence of the mass fraction on this ratio ($(\alpha/f)^{1/3}$) means that systems comparable to those of Jupiter, Saturn, and Uranus result for $10^{-6} < \alpha/f < 5 \times 10^{-4}$. Neptune is excluded because the capture of the large satellite Triton disrupted the original system. The application of the theory to Uranus requires additional processes that lead to the large obliquity after the satellite system was formed. An impact that tilted Uranus would disrupt the Canup-and-Ward-formed satellite system leading to mutual collisions among the satellites and reassembly in the new equator plane, or if an adiabatic process can be found to tilt Uranus, the satellite system would remain equatorial during the process (see Section 10.14.6). Both α and f are most likely to vary both in time and from planet to planet given the difference in the planet compositions and the likely decrease in f from solar composition as the solar nebula is evaporated. But values of α consistent with the above range of α/f and values of f equal to or less than that for solar composition are reasonable given the constant stirring of the disk by the continued influx.

To check these conclusions, the satellite formation is simulated by the N-body SyMBA code (Duncan *et al.*, 1998) based on the algebraic mapping integration scheme of Wisdom and Holman, modified to include the interaction of the masses with the gas disk and ongoing mass inflow (Canup and Ward, 2006). The satellites interact gravitationally with Jupiter and with each other. The background disk is treated analytically and interactions of the satellites with it are treated as small perturbations of the orbits. The radial surface density profile is approximated by $\sigma_g(r) = \sigma_0 (R_J/r)^{\gamma_g}$, and $H/r = c_s/r \Omega_k = b_0 (r/R_J)^{\gamma_c}$ gives the radial variation of the scale height. The scale height depends on the radial temperature distribution determined by the opacity. Values of $\gamma_g = 0.75$, $\gamma_c = 0.13$ and $b_0 = 0.07$ are estimated from the steady-state disk model of Canup and Ward (2002) with a constant opacity (see Appendix 1). Satellites experience a drag force because the gas is partially supported by the radial pressure gradient discussed in Appendix 1, where the satellites are given kicks at each time step proportional to the area to mass ratio in a direction opposite to the relative velocity between the satellite and the gas. The gas drag is only important while the satellites are small, where Type I interactions dominate the reduction in

eccentricity, orbital inclinations, and semimajor axis for larger satellites. After Papaloizou and Larwood (2000), Type I interactions are simulated with added accelerations $\mathbf{a}_I = -\mathbf{v}/\tau_I$, and $\mathbf{a}_{\text{damp}} = -2(\mathbf{v} \cdot \mathbf{r})\mathbf{r}/(r^2\tau_e) - 2(\mathbf{v} \cdot \hat{\mathbf{k}})\hat{\mathbf{k}}/(r^2\tau_i)$, where $\hat{\mathbf{k}}$ is a unit vector perpendicular to the plane of the disk, \mathbf{v} is the satellite velocity. The timescale τ_I for the Type I reduction of the semimajor axis is related to the timescale $\tau_e \approx 1.4\tau_i$ for the damping of the eccentricity and inclination by (Tanaka and Ward, 2004)

$$\tau_I = \frac{1}{C\alpha\Omega_K} \left(\frac{M_J}{m_s} \right) \left(\frac{M_J}{r^2\sigma_g} \right) \left(\frac{H}{r} \right)^2 = \frac{C_e}{C\alpha} \frac{\tau_e}{(H/r)^2} \quad [7]$$

where C_α and C_e are constants of order unity.

The solid inflow is represented by adding orbiting masses with random positions for $r < r_c$ at a rate proportional to F_{in}/f . Constraints on CPU time for the calculations force the added objects to be 50–600 km in radius – far larger than the 1 m-sized objects expected to follow the gas during the inflow. But once in orbit, the rapid accretion times imply that the particles would coagulate into sizes that are decoupled from the gas in times less than the viscous spreading time of the gas and to accumulate into larger objects on time-scales short compared to gas drag timescale τ_{GD} . Hence, starting with 50–600 km objects to accommodate the computational constraints should have little effect on the outcomes. Collisions within the disk are treated as inelastic mergers.

Since the satellites continue to accrete objects after they reach mass m_{crit} , their migration into Jupiter is hastened, and the value of M_T/M_J given by eqn [6] is an upper bound. This is illustrated in Figure 1 for a constant and uniform flux density F_{in} ($\gamma_{\text{in}}=0$), $\tau_G=5 \times 10^6$ years, $r_c=30R_J$, and $\alpha/f=10^{-6}, 5 \times 10^{-5}, 5 \times 10^{-4}$. The solid lines indicate the oscillating values of M_T/M_J resulting from the occasional loss of a satellite into Jupiter as a function of time normalized by the accretion timescale, and the dotted lines are the values of M_T/M_J from eqn [6], both for the indicated values of α/f . Generally the total mass of solids delivered to the disk will exceed M_T , so that several satellite systems with $M_T/M_J=2 \times 10^{-4}$ would be provided with the earlier ones being lost to Jupiter.

Figure 2 shows that for time-dependent influx rates with $\tau_{\text{in}} \sim 1.5 \times 10^6$ years, $r_c=30R_P$, $\gamma_{\text{in}}=0$ and $\alpha/f=6.5 \times 10^{-5}$ with $\alpha=6.5 \times 10^{-3}$, a system of satellites resembling the Galilean satellites is produced with $M_T/M_J \sim 3 \times 10^{-4}$. Recall that M_T/M_J is almost independent of τ_G (eqn [6]), but that disk

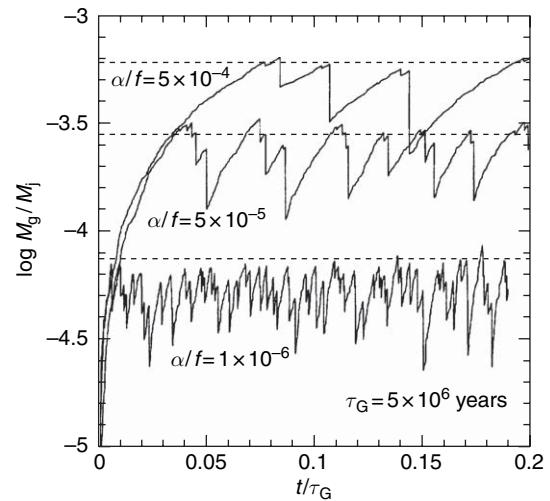


Figure 1 Total mass of satellites as a function of time for a constant rate of inflow for three values of α/f . Reprinted by permission from Macmillan Publishers Ltd: (Nature) (Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441: 834–839), copyright (2006).

temperatures are affected by F_{in} (with effective disk temperature $T_{\text{eff}} \propto F_{\text{in}}^{1/4}$ (Canup and Ward, 2002)). Early in the simulation when the influx is high, the masses of the satellites are given by eqns [5] and [6], but the satellites accreted will be rocky. The properties of the disk must evolve to the point where the midplane temperature of the disk is below 200 K at least beyond the point in the disk where Ganymede accretes a significant fraction of its mass, and the accretion rate of at least Callisto has a timescale greater than 10^5 years to accommodate the latter's possible lack of full differentiation. This timescale requirement persists even if Callisto turns out to be differentiated (McKinnon, 1997; Peale, 1999), since influx to the disk and hence the maximum accretion rate is limited by the low-disk-temperature constraint. As the influx wanes, the disk temperature is expected to drop and the component of ice in the satellites is increased. A final generation of Jovian satellites from solar composition inflow with $\tau_{\text{in}}=10^6$ years would form within a disk having temperatures below 200 K exterior to $15R_J$ with $\alpha=a \text{ few} \times 10^3$, a disk opacity $\kappa=O(0.1) \text{ cm}^2 \text{ g}^{-1}$, and a planet temperature $T_J=500$ K (Canup and Ward, 2002). The constraint of the outer satellites having a large icy component is thereby satisfied. The corresponding timescale for the accretion of Callisto at this stage is $>10^5$ years which is consistent with only partial differentiation according to the

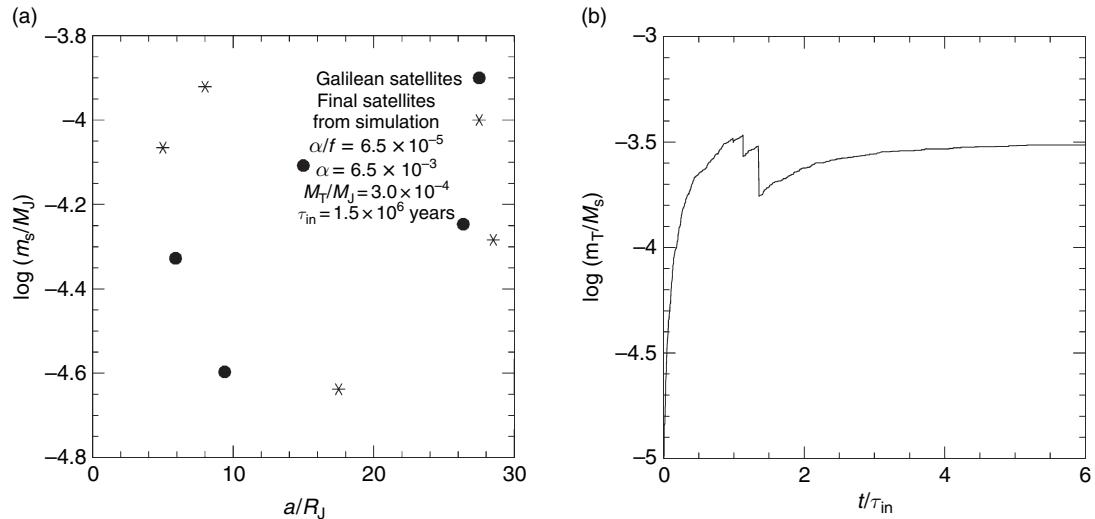


Figure 2 An example of a Jupiter-like system for a waning influx of material onto the disk. (a) Comparison of the masses and locations of the real Galilean satellites with the final set of satellites of a Jupiter-like system from a simulation of a waning influx of material. (b) The time dependence of M_T/M_J for the simulation resulting in the final set of satellites shown in (a). Reprinted by permission from Macmillan Publishers Ltd: (Nature) (Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441: 834–839), copyright (2006).

criteria of Stevenson *et al.* (1986). Neither the ratio of the total mass of the satellites to the planetary mass, M_T/M_J nor the distribution of the final satellites are affected much for $-1 < \gamma_{in} < 1.8$, so the results are given in Canup and Ward for $\gamma_{in}=0$. However, whether most of the mass is deposited close to the planet or far from the planet should affect the distribution of the ice fraction, but this effect of γ_{in} has not been extensively explored. (Canup, private communication, 2006).

In reality, the satellites are likely to be trapped into orbital resonances sometimes through differential migration of the forming satellites. Such captures were common in the integrations of Canup and Ward (2006) when there were larger outer satellites catching up with smaller inner satellites through Type I drift (Appendix 1) (Canup, private communication, 2006). Peale and Lee (2002) demonstrate an assembly of the Galilean satellites into the set of 2:1 mean motion resonances that comprise the present Laplace relation, where Ganymede is assumed to migrate on the 10^5 year timescale of Canup and Ward, and Io and Europa migrate on timescales that are double this value. All the current properties of the resonant system are produced in this scenario, which also requires significant tidal torque from Jupiter to prevent disintegration of the resonant structure from dissipation in Io. However, the assumption of the relative migration rates is *ad hoc*, and it remains to

be shown that detailed plausible disk properties including the opacity as a function of temperature and thereby position in the disk (Appendix 1), assumed viscosity, etc. can lead to relative migration rates that are consistent with the hypothesis. Still, during the satellite accretion, it is certain that inner small satellites will be trapped into orbital resonances with outer larger satellites that are migrating faster. Such trapping in resonances will hasten the demise of the small satellites interior to larger satellites, where such effects are naturally accounted for in the simulations of the mutual gravitational interactions during differential migration.

The Canup–Ward model successfully produces systems similar to the Galilean satellites (Figure 2). Its virtues include the following: (1) The model is formulated within a plausible model of a waning accretion of the planets, not only naturally producing the observed satellite–planet mass ratios for Jupiter, but also those for Saturn and Uranus (but see Sections 10.14.5 and 10.14.6 for possible caveats). (2) The waning accretion causes the disk to evolve naturally to sufficiently low disk temperatures to allow water ice condensation and incorporation of that ice into the satellites. (3) The main conclusions are insensitive to the inflow rate and only weakly dependent on the notoriously uncertain value of α . (4) The last generation of Jupiter satellites accrete sufficiently slowly to allow possible incomplete differentiation of

Callisto. (The latter constraint could be relaxed somewhat if it turns out that Callisto is in fact completely differentiated (McKinnon, 1997; Peale, 1999).) (5) The theory is clearly formulated in terms of parameters whose values can be adjusted and the consequences ascertained as new observations and theory provide more constraints. (6) Numerical simulations are consistent with the analytic approximations used in estimating the critical satellite mass attained before loss to the planet and for the total mass fraction M_T/M_J . (7) The simulations produce satellite systems similar to those observed for Jupiter, Saturn, and Uranus for plausible values of the parameters. (8) The values of α (and α/f) leading to the satellite systems similar to those observed are plausible given the turbulence that is likely to be generated by the continuing influx.

Refinements and further experimentation with the model could address the following questions: (1) What are the consequences of a more detailed treatment of the disk opacity accounting for the nature and quantity of dust as a function of radial distance? The dust opacity is likely to be maintained through mutual collisions during the accretion process. (2) What are the consequences of orbital resonances among the satellites during their formation and migration? (3) Although the disk was probably shadowed from the Sun by the solar nebula, could magneto-hydrodynamic turbulence resulting from ionization by cosmic rays or ultraviolet radiation from nearby stars constrain the values of α ? (4) What are the effects on disk dynamics of significant numbers of larger planetesimals that do not follow the inflowing gas? For example, could there be significant breakup of the forming satellites as the large planetesimals come crashing through the disk?

In the model of Alibert *et al.* (2005b) the flow of material through the circumplanetary α (kinematic viscosity $\nu = \alpha c_s^2/\Omega_K$) disk (Appendix 1) is determined from analytical fits (Veras and Armitage, 2004) of the ratio of the decaying accretion mass flux through the gap to the mass flux that would have occurred in an unperturbed solar nebula, which ratio is obtained from 2-D simulations of Lubow *et al.* (1999) and D'Angelo *et al.* (2002). The mass flux from the unperturbed nebula is taken from Alibert *et al.* (2005a). The mass is assumed to be added to the outer edge of the disk at $150R_J$, and all mass accreted by Jupiter passes through the disk. The mass flux is assumed to be the same at all radii in the disk while the solar nebula is still adding mass to the disk (phase 1). The vertical disk structure is determined by the

method of Papaloizou and Terquem (1999) involving the viscous dissipation according to the α model and the energy balance equation discussed in Appendix 1. The midplane pressure and temperature are thereby determined as a function of r and t constrained by the mass flux and not assigned an *a priori* radial dependence. Type I drift is applied to the accreting bodies until the Hill sphere of the satellite is comparable with the scale height after which type II drift is assumed to apply.

Satellites are assumed to grow linearly on various timescales from 10^4 to 10^6 years. Migration during growth determines where a satellite can be with a certain ice content for various states of the evolving disk, which thins and cools as the mass flux is reduced. For a given mass flux through the disk, a low value of α implies a higher σ_g . Since type I migration goes as the surface density, the migration is faster for low α disks. Since more massive disks are thicker, type I drift will persist for higher mass objects. If type II migration does prevail, its proportionality to α means it will last longer, so the distance a satellite will migrate is longer in a low α disk. At the beginning of phase 1, with a thick disk, temperatures are too high for icy particles to survive, and full type I drift results in the loss of all rocky satellites. Starting the formation of the rocky satellites Io and Europa toward the end of phase 1 (i.e., toward the end of mass accretion) allows their survival with full strength type I migration, near their current orbital radii if $\alpha > 2 \times 10^{-4}$, and the assumed timescale for accretion is $\tau_S < 10^5$ years. In phase 2 of the disk history, the accretion has ended, and the disk spreads and cools as the gas continues to accrete onto Jupiter. The formation of Callisto during this second phase when the disk has cooled sufficiently to have the ices of CO_2 and NH_3 or compounds of the latter to be stable, requires $\alpha \lesssim 10^{-3}$ to prevent the disk from dissipating too fast and an assumed formation time $\tau_S < 10^6$ years.

The assumption of mass deposition only on the outside of the disk is not consistent with the nature of the mass flux onto the disk obtained in the 3-D simulations of Bate *et al.* (2003). These latter calculations have superseded the earlier 2-D calculations, which were used to determine the mass flux to the disk. The authors imply that their disk is starved, but the mass content of the disk as a function of time is not given. Presumably the disk is indeed starved at the beginning of phase 2. But if this so, is there enough solid material remaining in the disk to make Callisto and Ganymede during this phase? One would need a minimum mass disk unless much of

the accretion of these two satellites occurred before phase 2, but in this case the disk is still too warm to allow the ices to exist. The paper lacks important details about the accretion process itself. For given disk parameters, will the timescales for the accretion actually be close to those assumed? This and other details such as the disk mass as a function of time, the nature of the transition between phase 1 and phase 2, and its timescale are needed to support the conclusions. Although the Alibert *et al.* model is placed in the context of the accreting planet during phase 1, in phase 2, the accretion has stopped and the model adopts the features of the closed MMSN for the formation of Callisto. Accretion timescales are assumed and so the formation scheme for Callisto is *ad hoc* and not constrained by the relevant timescales for spreading of the disk and the accretion times therein after accretion has stopped.

The Canup–Ward model (2002, 2006) in contrast is a more complete and plausible model of satellite formation within a model of the latter stages of the accretion of Jupiter. It is consistent with the composition and positional constraints and produces the gross features of the satellite systems of the giant planets for plausible choices of the controlling parameters, which differ from planet to planet. Some choices of the parameters yield systems comparable with the observed Galilean satellites. The vertical accretion onto the disk is consistent with the high resolution 3-D simulations, although the simulations lack sufficient resolution to constrain the radial dependence of the accretion. The accretion of the satellites in the circumplanetary disk is specifically treated as a function of disk properties that change as the accretion winds down, where plausible timescales for the various processes are assigned that are consistent with the observed properties of the satellites. The conclusions from the model are rather weakly dependent on unknown values of some parameters. Refinement of the model should be straightforward as constraints on the mass flux density distribution from higher resolution 3-D simulations become available and more sophisticated thermal modeling of disk structure is invoked.

10.14.5 Saturn System

The Saturn system is far different from the Jupiter system, and aside from the conclusion that the regular satellites accreted in a cold, dissipative disk, the details of the formation of the satellites therein are less understood. The densities of the satellites

Mimas, Enceladus, Tethys, Dione, Rhea, Titan, Hyperion, Iapetus are 1.15, 1.61, 0.96, 1.47, 1.23, 1.88, 0.6, 1.09 g cm^{-3} respectively, which imply a large ice content for all the satellites with perhaps considerable porosity in some. Rock-like mass fractions peak at Enceladus, Dione, and Titan giving a strikingly nonmonotonic density distribution. The mass fractions of rock in Enceladus, Dione, and Titan are similar to those of Ganymede and Callisto, where the higher density of Enceladus compared to Dione may reflect significant loss of water from the observed plumes at the South Pole (Porco *et al.*, 2006). The mole fraction of methane in Titan's atmosphere is about 5% in the lower atmosphere and the rest is essentially all N_2 except for traces of argon and small amounts of heavier hydrocarbons (Niemann *et al.*, 2005). Aside from a mole fraction of about 0.00003% for ^{36}A , there were no other primordial inert gases detected. Heavier hydrocarbons are thought to be products of photochemistry initiated by photodissociation of methane (e.g., Niemann *et al.*, 2005). The relatively short timescale for the survival of CH_4 from photodissociation of 10–100 My (Yung *et al.*, 1984; Wilson and Atreya, 2004), together with the lack of any isotope fractionation in the carbon of CH_4 like that seen in nitrogen and oxygen, means that CH_4 must be continuously supplied from Titan's interior (Nieman *et al.*, 2005). The satellites inside Titan's orbit comprise only 3% of the total mass of the satellite system. This small mass, the lack of any rock-dominated satellites (inferred from their densities) inside Titan's orbit, and the nonmonotonic distribution of densities comprise the major mysteries of the formation of the Saturn satellite system.

Titan is about 50% water as deduced from its mean density. If the N_2 in Titan's atmosphere were trapped in the icy planetesimals that went into Titan as direct condensation from the gaseous phase, or if it were trapped in amorphous ice or as a clathrate hydrate, the noble gases would have also been trapped. The low upper bound on the noble gases implies then that the N_2 came into Titan in the form of NH_3 (Niemann *et al.*, 2005, and references therein). One possibility is that the icy planetesimals formed at temperatures $> 75 \text{ K}$ so that the noble gases and CH_4 and were not captured, but NH_3 and CO_2 were. This scenario may require the continuous supply of CH_4 to Titan's atmosphere to be created through chemistry within Titan (Zolotov *et al.*, 2005). The constraints on the nebular conditions, especially the temperature, inferred from the constituents of Titan's atmosphere are just beginning to be explored

(e.g., Niemann *et al.*, 2005). We are certainly left with the constraint that the temperature eventually becomes cool enough in the Saturn disk to allow water ice to persist, but perhaps not so cold as to permit stable methane clathrates. But what happened to the rock inside Titan's orbit, and why are there no large satellites there?

Most disk models have surface density, mid-plane temperature and pressure increasing with decreasing r as in the Jupiter examples above. This normally would favor the formation of high-density rock-like major satellites close to the planet with icy satellites further out. That the opposite is true at Saturn means an evolutionary change in the disk properties must be responsible, and it seems difficult to effect such an evolution while the disk is still accreting, even in a starved disk scenario. This is contrary to our assertions for the Jupiter case where we argued for formation of the satellites during the final stages of accretion. For example, one scheme to account for the icy close satellites of Saturn is the maintenance of a disk temperature that is too high to allow the condensation of water vapor into ice particles inside the orbit of Titan, while all of the condensed silicate embryos spiral into Saturn. Later, the depleted, water-dominated disk cools to the point where icy embryos form, since ice is the only solid material available. Mosqueira and Estrada (2003a, 2003b) propose this scheme as one of several possible, but the discussion is inconclusive in the selection of one over another. For example, a higher turbulence (large α) is proposed in one case to dissipate the disk more rapidly to prevent formation of large satellites, but a small turbulence in another case makes the disk last longer and allows the loss of the silicate satellites leaving a water-rich disk to cool and form icy embryos. These authors prefer the action of a fully formed Titan on the inner disk, such as its removal on a 10^4 – 10^5 year timescale before large satellites could form. But the discussion of large number of alternative processes with different choices of poorly constrained parameters in each leaves a reader floundering for something firm to grasp in a Titan influence on the formation of the inner satellites.

The Mosqueira–Estrada model for the formation of the Saturn satellites starts with the minimum mass nebula based on the total satellite masses, but not on the distribution of mass in the satellite system. The temperature and surface mass densities vary as $1/r$ as in the Jupiter model, with $T = 100\text{ K}$ at Titan's distance based on the assumption that Titan's methane came from methane ice, which is contrary to the

results of Niemann *et al.* (2005) and Zolotov *et al.* (2005) based on Titan's atmospheric chemistry. Their closed disk may be relevant to the late stage formation of the inner icy satellites, when accretion is likely to have stopped, and the disk cools after having lost its silicate embryos. As in their Jupiter model, there is a steep surface density gradient just outside the centrifugal radius defined above, which for Saturn, is just outside the orbit of Titan. Interestingly, the capture of Hyperion in the 4:3 mean motion resonance with Titan seems to require such a steep density gradient ($\sigma_g \sim r^{-3}$) (Lee and Peale, 2000). Hyperion was most likely accreted in the region outside Titan as Titan itself formed, where gas drag is necessary to bring Hyperion into the 4:3 resonance. The accretion is essentially independent of the presence of the orbital resonances since the mutual potentials in close encounters of the satellitesimals exceed the depth of the resonance potentials in magnitude. As Hyperion nears its final mass, probably considerably larger than the current mass because pieces chipped off Hyperion typically do not remain in the resonance and are not re-accreted, gas drag overwhelmingly selects the 3:2 mean motion resonance preventing migration to the 4:3 if the surface mass density decreases less rapidly than r^{-3} (Lee and Peale, 2000). Mosqueira and Estrada form icy Iapetus in their outer, low surface mass density disk, where it migrates by type I drift to the point where this torque vanishes because of the steep density gradient between inner and outer disks. But migration of Iapetus, with diameter $\sim 1530\text{ km}$, is only marginally dominated by type I torques versus those from gas drag.

Alibert and Mousis (2006) extend their model of the Jupiter subnebula to Saturn. This is an attempt to model the formation of the Saturn satellites within a model for the formation of Saturn itself, but only the formation of Titan is considered. Again the model suffers because the accretion of material on the outer edge of a steady state disk with declining mass flux does not agree with the results of Bate *et al.* (2003) whose 3-D simulations find mass raining on the top and bottom of the disk. Ices are only stable during phase 2 when the solar nebula has disappeared and the system becomes closed with a cooling and spreading disk. This means Titan forms from a closed disk with sufficient mass to accommodate Titan and the other satellites plus that lost from the disk to Saturn. The details of the transition between phase 1 (with accretion) and phase 2 (without accretion) are missing in the Jupiter application and here. A key

unsupported assumption is that the timescale for the disappearance of the solar nebula is short compared with the timescale for formation of the satellites within and the disappearance of the subnebulae, which assumption is contrary to the short disk timescales inferred in Appendix 1 or those controlled by gas drag in the Mousqueira and Estrada models. The Alibert and Mousis model does not address the formation of the icy satellites inside Titan.

The Canup–Ward model (2002, 2006) described in Section 10.14.4 has the virtue that it is simultaneously applicable to any planet with the substitution of the appropriate parameters. Their model is constructed in the context of the waning accretion of Saturn, and it is more complete and of higher plausibility than the model of Alibert *et al.* (2005b). Accretion of the satellites is simulated with an N-body code described in Section 10.14.4, while the satellites migrate due to gas drag and Type I drift. Jupiter has a system dominated by four large satellites, whereas Saturn has one large satellite and numerous small satellites. Both distributions can result from the cycles of satellite accretion and loss depending on the timing of the mass fraction oscillations see in Figure 1. The systems cycle through states where there are only one or two of the largest satellites left accompanied by several smaller

satellites that have formed in regions vacated by larger satellites with the cycle repeating until the inflow ceases. The distribution of satellite masses at the end of a simulation depends otherwise on the choice of α and f . Figure 3a compares the distribution of Saturn's satellites with a distribution of satellites obtained from a simulation with $r_c = 30R_S$, $\gamma_{in} = 0$, $\alpha/f = 6 \times 10^{-5}$, and $\alpha = 6 \times 10^{-3}$. A seven satellite system has resulted with $M_T/M_S = 1.8 \times 10^{-4}$. Figure 3(b) shows the evolution of M_T/M_S resulting in the distribution of satellites in Figure 3(a). Notice that there is the loss of a large mass satellite just before accretion has almost stopped leaving the smaller interior satellites in its place. The most massive satellite at $\sim 14.6R_S$ contains 70% of the mass of the total satellite system mass and its close orbital spacing to the satellite at $11.3R_S$ suggests a future collision may occur. There are two satellites bracketing the position of Iapetus but with larger orbital eccentricities. The inner satellites in this simulation are sufficiently close together that collisions are likely with a reduction in the total number of satellites.

Canup and Ward point out that near the end of accretion onto the disk around Saturn, the ice line moves to within a few Saturn radii allowing the accumulation of icy satellites. However, even though the inner satellites shown in Figure 3 could have a

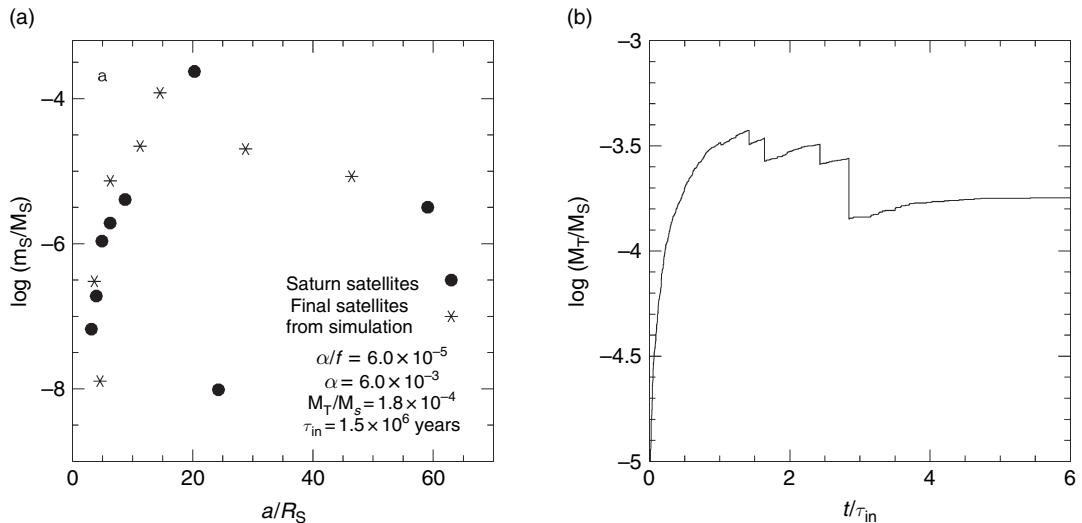


Figure 3 An example of a Saturn-like system for a waning influx of material onto the disk. (a) Comparison of the masses and locations of the real Saturnian satellites with the final set of satellites of a Saturn-like system from a simulation of a waning influx of material. (b) The time dependence of M_T/M_S for the simulation resulting in the final set of satellites shown in (a). Note in (b) the loss of a large satellite late in the evolution allowing the small inner satellites to survive. Reprinted by permission from Macmillan Publishers Ltd: (Nature) (Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441: 834–839), copyright (2006).

large component of ice, the apparently large increase in the fraction that is ice from say Titan to the inner satellites remains unexplained. We can perhaps account for the replacement of a large satellite inside Titan's orbit with small satellites within the Canup–Ward scenario, but what happened to the rock? Perhaps a more daunting puzzle is the nonmonotonic distribution of densities that imply a similar non-monotonic distribution of the rock fraction.

Iapetus presents a special problem in that its orbit is nearly circular, but inclined by about 7.49° relative to the local Laplacian plane. The latter inclination would imply a capture origin that is inconsistent with small orbital eccentricity ($e_I = 0.0283$). The Laplacian plane is that plane on which a satellite's orbit precesses at nearly a constant rate with nearly a constant orbital inclination. The Laplace plane nearly coincides with the equator plane for close satellites due to torque from the planetary oblateness, but it coincides nearly with the planet's orbit plane for distant satellites due to the solar torque. An extensive debris disk settling to this plane would therefore be warped as the local Laplace plane changed from equator plane to orbit plane with distance from the planet (e.g., Ward, 1981). The regular satellites accumulated from debris disks at distances where the oblateness of the planet dominated the location of the Laplace plane. But at Iapetus' distance from Saturn, the local Laplace plane is inclined by 15° to Saturn's equator, where the ascending nodes on the orbit plane of Saturn's equator (obliquity = 26.7°) and the Laplace plane coincide. Can Iapetus have accumulated in the debris disk as implied by its small orbital eccentricity, yet be inclined to the local Laplace plane by 7.49° today?

Ward (1981) offers a tidy affirmative answer by pointing out that the debris disk itself contributes to the precession of objects within it. The Laplace plane thereby would have a lower inclination to Saturn's equator plane at the time the disk was present, and Iapetus could have accreted there as assumed in the above discussion. Iapetus' orbit plane precesses with a current period of about 3000 years (Ward, 1981), and this period would have been shorter when the disk was present. If the disk could be dispersed on a timescale short compared to the orbit precession period at the time, the Laplace plane would rotate toward its current position and leave Iapetus with its observed low eccentricity and 7.49° orbital inclination. The short dispersal time necessary, $O(100)$ years, is not consistent with accretion times of the satellites of order 10^3 – 10^4 years in the theories discussed above, or for allowing sufficient time for rocky satellites to spiral

into Saturn in the inner disk leading to the formation of icy satellites there. Viscous dissipation of the disk on this short timescale requires values of $\alpha = O(1)$ (Ward, 1981), which is implausible. Because of the difficulties imposed by such rapid disk dispersal, the anomalous orbit of Iapetus is still a puzzle. A capture origin seems unlikely, because sufficient gas drag to circularize the orbit would also have damped the inclination to the local Laplace plane.

The Canup–Ward model is perhaps a first step toward an understanding of the origin of Saturn's satellite system, since a system with no large satellite interior to Titan seems attainable. One caveat for the application of the Canup–Ward model to Saturn is that Saturn probably did not open a gap in the nebula, so the disk may have received its mass in a way quite different from the scenario we described for Jupiter that was supported by the Bate *et al.* (2003) simulation. A different accretion disk model may have to be formulated for those planets that do not open gaps in the solar nebula. On the other hand, as the solar nebula is dissipated, it becomes optically thin and may cool to the point that Saturn could open a gap during the last stages of its accretion. The loss of the silicate embryos while the nebula was too warm to allow water ice condensation followed by cooling to allow the remnant ice to accrete onto the small satellites is an attractive alternative (Mosqueira and Estrada, 2003a, 2003b). But there have been no detailed models of disk evolution with the appropriate distribution of timescales that would maintain the disk while keeping temperatures high long enough to allow migratory loss of the rocky satellites, and still maintain the disk while it cooled to allow ice dominated satellites to accrete. At the same time Titan must be prevented from spiraling in while all of this takes place, and Iapetus must not drift into an orbital resonance with Titan. There has also been no convincing post-formation scenario proposed that could reshuffle the distribution of densities and inferred rock fractions to produce the observed characteristics of the inner satellites. We are much closer to understanding the details of the origin of Jupiter's satellites than we are to understanding the origin of the Saturn system.

10.14.6 Uranus System

The Uranian system of satellites is Jupiter-like in the sense that it contains a few large satellites with $M_T/M_U \approx 1 \times 10^{-4}$. The densities/radii of Miranda,

Ariel, Umbriel, Titania, and Oberon are 1.2/236, 1.7/579, 1.4/585, 1.7/789, 1.6/761 ($\text{g cm}^{-3} \text{ km}^{-1}$), respectively. The Canup–Ward (2006) model can produce a Uranus-like system of satellites, where the lower rate of inflow and cooler disk temperatures about the smaller mass planet produce satellites with a mixture of rock and ice that is consistent with the densities. However, these authors note the difficulty of moving the disk material from heliocentric orbit to the equatorial plane of Uranus when the obliquity of Uranus is 97° . On the other hand, if the Uranus obliquity were small when the satellites formed, and only later tipped over by an adiabatic process, the Canup–Ward model would be immediately applicable. But with more certainty than the Saturn case, Uranus will not open a gap in the solar nebula, and the influx may have a different distribution over the disk than that appropriate to Jupiter. So the Canup–Ward model may have to be modified. But investigating the consequences of modifying the distribution of the inflowing of material, for example, would seem to be straightforward given the clear parameterization of the theory.

A means of increasing the Uranian obliquity about 500 My after the formation of the planets and their satellite systems has been proposed by Brunini (2006a). The Brunini model is based on a recent proposal for the establishment of the architecture of the outer solar system (Tsiganis *et al.*, 2005). The latter proposal envisages an initially compact solar system where all of the major planets were inside 17 AU with Jupiter and Saturn near their current distances from the Sun but with a separation that placed them closer than that corresponding to the 2:1 orbital resonance. The gaseous solar nebula was dispersed, but a sea of planetesimals remained mostly beyond the furthest ice giant. From the scattering of the few planetesimals remaining in orbits crossing those of the major planets and those scattered in from the region beyond the furthest ice giant, Jupiter moved closer to the Sun and Saturn further away until they crossed the 2:1 orbital resonance, where the diverging orbits prevented capture into the resonance (e.g., Peale, 1986). However, Saturn's eccentricity is excited by the passage, and its perturbations of Uranus and Neptune caused them to evolve into crossing orbits with each other and with Saturn. There were thereby many close encounters among Neptune, Uranus, and Saturn that scattered the ice giants out to the far reaches of the solar system, stirred up the sea of planetesimals whose scattering damped the inclinations and eccentricities

of the three outer planets through dynamical friction while causing them to migrate further away from the Sun. Numerical simulations demonstrate that it is possible to attain the current distribution of the major planets by this scheme (Tsiganis *et al.*, 2005). Brunini uses many close encounters between Saturn, Uranus, and Neptune to change the obliquities of the planets, where he finds a distribution of obliquities at the end of the simulation that match the observed obliquities reasonably well, including that of Uranus.

However, after being shown analytic and heuristic arguments that refute the claim of large obliquity changes from the encounters coupled with numerical verification of these arguments, Brunini (2006b) has withdrawn his paper. Details of the demise of the Brunini hypothesis are found in Lee *et al.* (2006, 2007), and there remains no accepted means of changing Uranus's obliquity adiabatically while maintaining its equatorial system of satellites.

We are left with the hypothesis that the Uranian obliquity is due to a giant impact during the last stage of accretion (Slattery *et al.*, 1992). For this scenario, the regular satellites of Uranus might have formed from an equatorial disk that was created by the impact (Stevenson, 1984; Slattery *et al.*, 1992). In contrast to the opinion expressed in Section 10.14.4 against starting with a pristine disk for the formation of the regular satellites of Jupiter, here is an example of the creation of a disk around a giant planet with perhaps 1–3% of the mass of Uranus (Slattery *et al.*, 1992) from which the satellites containing about 10^{-4} of the Uranus mass could be formed. If the solar nebula had already been removed at the time of impact, there would be no further accretion after the disk was formed. Considerable viscous spreading would be necessary while the disk is hot and opaque to spread an initially compact disk out beyond $25R_U$ where the satellite Oberon, now at $22.8R_U$, could form (Canup and Ward, 2000; Ward and Canup, 2003). Viscous dissipation during the spreading could keep the disk hot and possibly prevent premature accretion of the satellites. However, if the solar nebula had already been dispersed at the time of the giant impact, Uranus must have already possessed a satellite system. What happened to those satellites? Were they scattered into crossing orbits, shattered by mutual collisions, and incorporated into the new disk? If in fact most of the mass of the Uranian satellite system was reassembled in the new equator plane from the shattered masses of the preexisting system with only a minor contribution from the impact-generated disk, the necessity of a large

amount of viscous spreading would be reduced. The Canup–Ward result of $M_T/M_U \sim 10^{-4}$ could then account for the total mass of the system whether or not a giant impact occurred.

If we pursue the giant impact origin of the satellites, the distribution of densities is consistent with rocky cores accreting first as the already spread out disk cools and then adding icy mantles when the temperature became sufficiently low to allow the condensation of ice. The satellite radii are such that the gas drag timescale would dominate the migration of embryos that finally merge to make the final satellites. Type I migration would only dominate for the larger satellites after assembly was nearly complete. The lower density of Miranda implies a higher ice content or porosity than that of the other satellites, and if this trend continues to the small ring satellites inside Miranda’s orbit, it is consistent with the loss of rocky embryos to Uranus before the disk cooled sufficiently to allow ice particles to exist in the inner region. This would be the same hypothesis mentioned above for the inner Saturn satellites, but on a smaller scale.

Given an initially hot, compact disk following the giant impact, it would be a rather straightforward exercise to determine its viscous spreading as a function of the α parameter. The constraint on α necessary to spread the disk beyond the current orbital distance of Oberon together with self-consistent opacities would determine the temperature history and tell us if accretion could be delayed until the disk is mostly spread out (Ward and Canup, 2003). The mid-plane temperature history as a function of distance from Uranus coupled with the migration of the embryos would test the inference of increasing ice fraction close to the planet, where rocky embryos were plausibly lost to Uranus. The only constraint on the initial disk content, temperature, and radial extent comes from the SPH simulation of Slattery *et al.* (1992), where 1–3% of a Uranian mass of material ended up in orbit. However, this simulation contained only 5000 particles in the proto-Uranus and 3000 particles in a $2M_E$ impactor. Rather substantial changes in the outcome resulted from SPH calculations of the lunar-forming impact on the Earth as the resolution was increased substantially above that of earlier calculations (Canup, 2004b), so it is important to do simulations of the Uranus impact with the highest resolutions currently practical for reliable constraints on the initial disk properties.

At the time of this writing (Aug 2006) neither high-resolution simulations of a Uranus impact nor detailed analytic histories of the resulting equatorial disk had been published. The creation of the disk at the time of impact and formation of the satellites therein following a viscous spreading to cover the extent of the current regular satellite system is plausible on the surface. But the detailed calculations are necessary to insure that observations of the current system are matched with plausible assumptions about the expected processes that were operating with plausible choices of the parameter values and timescales therein. If the impact occurred after the nebula was dispersed, this model of the formation of the Uranian satellites, now incorporating a pre-existing satellite system, could be much cleaner and less complicated than that for either the Jupiter or Saturn systems because we would not have to worry about unknown accretion rates, which continuously decline, and other effects of the solar nebula. Although a plausible and detailed model of the formation of the Uranian satellites could emerge from further study of this system under the assumption that a giant impact tilted Uranus, the impact origin of the satellites remains uncertain. If a method of tilting Uranus at a slow enough rate that the satellites remain in their equatorial orbits turns out to be viable, then the formation of the satellite system as the accretion of Uranus wanes is the likely alternative. The fact that the Canup–Ward (2006) (Section 10.14.4) model produces Uranian-like systems for plausible values of the timescales and other parameters already provides a likely, albeit possibly modified scenario in this circumstance.

10.14.7 Neptune System

A plausible theory has emerged for the current configuration and content of the Neptune system of satellites because Triton’s destruction of the original system during its capture has erased all concern for accretion disks, gaps, solar nebula interactions, etc. The dominant characteristic of the Neptune satellite system is the existence of the large satellite Triton (NI) ($R = 1353$ km), in a close ($r \approx 14.3R_N$, ($R_N = 24\,766$ km)) circular, retrograde orbit (obliquity 156.8°). Neptune also has relatively few known satellites compared to the other major planets, and all but two of those, Triton and Nereid (NII), were unknown until the Voyager spacecraft observations (Smith *et al.*,

1989) and subsequent ground-based discoveries (Holman *et al.*, 2004). Nereid's orbital eccentricity of 0.75 brings it no closer than about 1.4×10^6 km from Neptune's center because of its extremely large semi-major axis of 5.51×10^6 km ($222.6R_N$) – well outside Triton's distance of 3.54×10^5 km ($14.3R_N$).

The massive satellite Triton is blamed for most of the features of this unusual system. The retrograde orbit means that Triton was almost certainly captured intact from heliocentric orbit. It has been proposed that Triton was captured in a three-body exchange where Pluto was expelled from orbit around Neptune (Farinella *et al.*, 1980), that it was captured by gas drag (McKinnon, 1984; McKinnon and Leith, 1995), or that it collided with a satellite already in orbit about Neptune with a mass a few percent of Triton's mass (Goldreich *et al.*, 1989). The first hypothesis suffers from the likely multiple subsequent close encounters of the ejected Pluto with Neptune and essentially no means of leaving Pluto in the currently stable 3:2 mean motion resonance with Neptune, whereas the second is perhaps less probable than the collisional capture.

A collision would have been sufficient to capture Triton into a very eccentric orbit extending a significant fraction toward the Hill sphere boundary at $r_H \approx (m_N/3m_\odot)^{1/3} a_N \approx 4.5 \times 10^3 R_N$ (a_N = heliocentric distance) while not destroying it. However, an alternative method of capture of Triton involving the disruption of a binary KBO with characteristics similar to the Pluto–Charon binary pair has been proposed (Agnor and Hamilton, 2004, 2006). Binary systems among the small bodies in the solar system appear to be ubiquitous with $\sim 16\%$ of near-Earth asteroids (Margot *et al.*, 2002), $\sim 2\%$ of main belt asteroids (Merline *et al.*, 2002), and $\sim 11\%$ of KBO (Stephens and Noll, 2006) having companions. The fraction of small bodies in binary pairs is probably larger than these such that binary–planet encounters are likely. When a binary pair comes sufficiently close to a planet such that the Hill sphere of the pair in the planetary gravitational field is smaller than their separation, they become tidally unbound. Depending on the phase of the relative motion in the binary and the inclination of the orbital plane with that of the encounter trajectory, one of the planetesimals can be left in bound orbit about the planet while the other escapes. A simulation with binary pair masses $m_1 = m_T$ and $m_2 = 0.1m_T$ separated by $20R_1$ in circular orbit, where m_T is the mass of Triton and R_1 is the radius of m_1 , closest approach distance $8R_N$ well inside the tidal disruption radius $r_{\text{td}} = 26R_N$ (R_N = Neptune radius), resulted in 95% of

the binaries being disrupted for $0 < v_\infty < 2.0 \text{ km s}^{-1}$, $\sim 50\%$ of the encounters resulting in capture of either m_1 or m_2 for $v_\infty \lesssim 0.35 \text{ km s}^{-1}$, and $\sim 50\%$ of the encounters resulting in capture of m_2 for $v_\infty \lesssim 0.35 \lesssim 1.55 \text{ km s}^{-1}$. For latter velocity interval, capture of m_1 is still possible but rare. Although either collisional capture (Goldreich *et al.*, 1989) or binary exchange capture (Agnor and Hamilton, 2004, 2006) can yield reasonable orbits that can damp to the observed retrograde circular orbit observed today, Agnor and Hamilton show that with at least 11% of KBOs being binary systems, capture by collision has negligibly small probability compared to the binary exchange capture. If the original satellite system of Neptune was similar to that of Uranus, collision of Triton with any satellites this large would have destroyed it making the currently inclined orbit of a recollect Triton impossible (Ćuk and Gladman, 2005). So it is almost certain that Triton was captured gently without such a risk, although a collisional capture remains an improbable possibility.

With $k_T (= k_2) = 0.1$ and $Q_T = 100$, Goldreich *et al.* (1989) find that an initial Triton orbit with semimajor axis $a_T \approx 10^3 R_N$ would damp to nearly its current circular orbit with $a_T \approx 14.3 R_N$ from tidal dissipation in Triton in about 4 to 5×10^8 years – comfortably shorter than the age of the solar system. However, this timescale for the decay of Triton's eccentricity must be at least two orders of magnitude too large. The large amount of dissipation probably melted a significant fraction of the satellite, which drastically hastens tidal evolution (Shock and McKinnon, 1993; McKinnon *et al.*, 1995). The Love number k_2 would then be $O(1)$ (closer to that of a fluid sphere) instead of $O(0.1)$ and $Q_T \sim O(10)$ instead of $O(100)$. In fact, if we substitute Triton's parameters into the second term on the right-hand side of eqn [58], the timescale for the eccentricity damping so obtained is $\tau_{eT} = 6.31 \times 10^4 Q_T/k_T$ years. The tides raised on Neptune by Triton will have a negligible effect (Goldreich and Soter, 1966; Banfield and Murray, 1992). With the Goldreich *et al.* choices of k_T and Q_T , $\tau_{eT} \sim 6 \times 10^7$ years, and several times this timescale required for complete damping approaches the several hundred million years damping time found by Goldreich *et al.* With the more likely value of k_T/Q_T this time is $< 10^7$ years.

However, tidal dissipation may have been less efficient in the circularization of Triton's original eccentric orbit than inferred by Goldreich *et al.* because solar perturbations cause periodic increases

in periapse distances (Benner and McKinnon, 1995). The Goldreich *et al.* analysis ignored these Kozai oscillations in eccentricity (see Section 10.14.9), which would have occurred if the original maximum orbital inclination was comparable to the current one. These 1000 year variations in e_T cause the mean periapse distance to be much larger than that obtained by Goldreich *et al.*, with the consequence that the tidal damping time, with the same values of Q_T and k_T and expression for dissipation used by Goldreich *et al.* would yield a timescale of 3.5 Gy for the eccentricity reduction (Cuk and Gladman, 2005). Reducing this number to 35 My with the more likely values of Q_T and k_T above still make the tides adequate, however. Although now unnecessary, another means of dissipating orbital energy and angular momentum might have been operating. An original satellite system like that of Uranus extending from about $5R_N$ to beyond $20R_N$, would have been perturbed by Triton into crossing orbits, and the satellite system would be thereby destroyed. During the times when Triton's periapse was inside $20R_N$, energy would be removed from Triton's orbit impulsively at each passage through the resulting disk of material leading to circularization times of 10^4 – 10^5 years while Triton cleared most of the material from the disk outside about $5R_N$ (Cuk and Gladman, 2005).

Consistent with the capture scenario and subsequent melting, Triton was found to have a young and active surface. The observed, currently active plumes are almost certainly solar driven (Smith *et al.*, 1989), but the geologically young volcanic plains and cantaloupe terrain are clearly endogenic, where such surface modification may be ongoing. (Shenk and Jackson, 1993; Stern and McKinnon, 2000). Triton's density is about 2 g cm^{-3} which indicates some ice content. Triton's original constituents were obtained directly from the solar nebula at its location of formation, but there may be significant contamination with debris from the original satellite system. The volatile content was probably reduced during the rather extreme tidal heating during the eccentricity damping, although some N_2 was retained from the early dense atmosphere and most of the water is still there (Lunine and Nolan, 1992). Triton and Pluto are similar in size and density, but Pluto may have lost a lot of volatiles from the impact that created Charon (see Section 10.14.8), and one cannot infer that this loss is the same as that from Triton. The current density does not constrain the place in the solar nebula where Triton may have been formed. But an initial composition similar to that of Halley's comet is consistent with

observations of the surface ices, where the severe reduction of CO relative to N_2 and CO_2 can be understood by hydrothermal processing resulting from the tidal heating (Shock and McKinnon, 1993).

While Triton's orbit was extended, any regular satellites inside about $5R_N$ would also have been disrupted and destroyed from the periodic dips of the periapse inside 5 to $8R_N$, depending on the initial semimajor axis. There are likely to have been several satellites inside the periapse distance of Triton by comparison with the regular systems of the other major planets, but the nature of this initial system, whether from an accretion disk or from a disk resulting from a giant impact, as inferred for Uranus, remains unknown. Orbit eccentricities would have been limited to maximum values near 0.3, since the rate of decrease of the eccentricity by tidal dissipation within the satellite exceeds Triton's ability to increase it for larger values (Banfield and Murray, 1992). Two satellites with semimajor axes of $3R_N$ and $5R_N$ would have overlapping orbits that would persist in overlapping for times at least of the order of the eccentricity damping time of $\sim 10^5$ – 10^7 years. A collision between any of Neptune's current inner satellites with its neighbor would lead to their mutual destruction. This follows from the relation (Stevenson *et al.*, 1986)

$$\frac{1}{2} m_i v_i^2 \sim m_s S + \frac{3}{5} \frac{Gm_s^2}{\gamma R_s} \quad [8]$$

where m_i and m_s are the masses of the impactor and satellite respectively, v_i is the relative velocity at impact, R_s is the satellite radius, $S \sim 10^6 \text{ erg g}^{-1}$ is the material binding energy and $\gamma \sim 0.1$ is a factor introduced to account for the inefficiency in converting the impact kinetic energy into kinetic energy of the fragments. The impactor kinetic energy must exceed the energy stored in material strength plus the self-gravitational energy by a sufficient amount to break up the body. From eqn [8] and eccentricities of 0.3, Naiad (NIII) could destroy all of the satellites except Proteus (NVIII), and any of the satellites NIV to NVII could destroy Proteus. This implies that the current satellite system could not have existed prior to Triton's orbit circularization (Banfield and Murray, 1992). The debris from the first generation of satellites would settle into the equatorial plane in circular orbits and re-collect into a second generation of inner satellites with nearly circular orbits and zero inclinations, where all memory of Triton's perturbations would be thereby lost. Had any of these inner satellites survived destruction, their orbits would have been left inclined to Neptune's equator plane. The lack of such inclined

orbits among the inner satellites must mean that there are no survivors among those early inner satellites.

The inner satellites observed today are not those formed after the circularization of Triton's orbit. The flux of comets through the system would have destroyed all of the current inner satellites except for Proteus (1989N1) over the lifetime of the solar system (Smith *et al.*, 1989; Colwell and Esposito, 1992). The debris from such a destructive collision would settle to the equatorial plane where the satellites can re-accrete. The rings of small particles are likely to be remnants of these collisional breakups. Any distant irregular satellites of Neptune (Holman *et al.*, 2004) would also have their orbits perturbed drastically by Triton if they were captured before or during Triton's capture and orbital decay and their numbers depleted (Ćuk and Gladman, 2005).

10.14.8 Pluto System

Pluto and its large satellite Charon orbit about their center of mass with a period of 6.3872304(11) days at a separation of 19 571 km (Buie *et al.*, 2006). The rotation of both bodies is synchronous with the orbital motion, which makes this system the only one in the solar system known to have reached the end point of tidal evolution. Like the Earth–Moon system, the Pluto system has a high specific angular momentum that favors the formation of Charon by a large oblique impact (McKinnon, 1984, 1989; Stern *et al.*, 1997). Canup (2005) has performed SPH simulations of impacts on the Pluto precursor with impactor mass comparable to the target mass with a variety of compositions and differentiation states. The most probable outcome of a simulation that is successful in placing a significant amount of mass in orbit outside the Roche distance is for Charon to be launched as an intact satellite in a highly eccentric orbit. In these examples, Charon is composed almost entirely of material from the impactor, which results from its being sheared off in the oblique impact. Much of the discussion of the giant impact creating the Moon in Section 10.14.2 is applicable to Pluto–Charon, but the impact associated with the latter involves much less energy. There is little heating of Charon, little water ice is vaporized, but there is probably sufficient energy deposited in Pluto at impact coupled with radiogenic heat to cause its differentiation (Canup, 2005).

Two new satellites of Pluto have just been discovered (Weaver *et al.*, 2006), and from a series of

HST images taken before the discovery, Buie *et al.* (2006) have determined that the orbits of the new satellites, originally called P2 and P1 but now named Nix (inside) and Hydra (outside), are nearly circular and coplanar with the orbit of Charon. Best-fit Kepler orbits to 12 composite data points by Buie *et al.* yield periods, semimajor axes, and eccentricities of 24.8562(13) d, 38.2065(14) d, 48 675(121) km, 64 780(88) km; 0.0023(21), 0.0052(11), respectively. The orbital periods of Charon, Nix, and Hydra are nearly in the ratios 1:4:6, but no resonance variables are likely to be librating at the present time (Lee and Peale, 2006). The new satellites orbit about the center of mass of Pluto and Charon, and that position from the orbit fitting procedure gives a mass ratio of 0.1165(55). Pluto's radius from the mutual event light curves is 1164(22.9) km (Young and Binzel, 1994) and Charon's radius from a stellar occultation is 603.6(1.4) km, (Sicardy *et al.*, 2006). With the masses determined by Buie *et al.*, these radii yield densities of 2.03(0.06) and 1.65(0.08) g cm⁻³ for Pluto and Charon, respectively. The physical parameters of Pluto and Charon are consistent with the results of the giant impact origin of the system (Canup, 2005). Lee and Peale (2006) show that the orbital motions of Nix and Hydra are markedly non-Keplerian because of large Charon–Pluto mass ratio. As the masses of the new satellites are increased within the range allowed by the uncertainty in the albedos, the close proximity of the 3:2 mean motion resonance between Nix and Hydra causes increasing fluctuations in the eccentricities. Although the maximum masses allowed within these constraints allow possible libration of the 3:2 resonance variable involving the periape longitude of Hydra, these maximum masses are probably already ruled out because they would also excite Charon's eccentricity beyond the current observational uncertainties (see Stern *et al.*, 2003). A dynamic fit to a sufficiently dense set of observational data will yield constraints on the masses of Nix and Hydra (Lee and Peale, 2006). The fact that Nix and Hydra are in orbits coplanar with Charon's orbit is convincing evidence that they were created during the same impact event that left Charon in bound orbit about Pluto (Stern *et al.*, 2006).

However, Nix and Hydra reside at \sim 42 and \sim 56 Pluto radii (R_P) from the Pluto–Charon center of mass, and debris from the impact is not likely to end up beyond $15R_P$ (Canup, 2004a, 2005). In addition to the nearly intact Charon that is the most likely outcome of a large, oblique impact, some amount of debris orbiting close to Charon's orbital plane both

inside and outside Charon's initial orbit will remain. Charon's orbit must expand from its initial size in the process of reaching the current state of completed tidal evolution of dual synchronous rotation (Dobrovolskis *et al.*, 1997). Ward and Canup (2006) demonstrate what may be the only viable scheme for accounting for the current location of Nix and Hydra if they are created in the large impact that formed Charon. One can infer from the closeness of Nix and Hydra to the 4:1 and 6:1 mean motion resonances with Charon that the satellites probably occupied these resonances in the past. Being so trapped would mean they could be pushed out ahead of Charon as Charon's orbit tidally expanded. But the resonance must be chosen with care for this to happen.

There are a number of resonances which can be occupied at the 4:1 and 6:1 mean motion commensurabilities characterized by the angular arguments of terms in the expansion of the perturbing potential

$$\phi_{ml} = (m + 1)\lambda - \lambda_c - (m - l)\varpi - l\varpi_c \quad [9]$$

where λ and λ_c are the mean longitudes of the satellite and Charon, ϖ and ϖ_c are the longitudes of periape. The angle ϕ_{ml} is written for m th order resonances of the form $m+1 : 1$ with $m=3$ for Nix and $m=5$ for Hydra. The terms in the expansion of the perturbing potential are represented by $\Phi_{ml}(a, a_c, e, e_c) \cos \phi_{ml}$, where terms involving orbital inclinations are ignored and where $\Phi_{ml} \sim e^{m-l} e_c^l$. The only resonances at the $m+1 : 1$ mean motion commensurability that can preserve the satellites during Charon's orbital expansion are for $m=l$ involving only the eccentricity of Charon's orbit in the coefficient Φ_{ml} . This follows from the fact that all those resonances involving e and hence ϖ lead to steady growth in e as the orbits expand with eventual instability. The resonance involving only the eccentricity and longitude of periape of Charon is called a co-rotation resonance, since the satellite orbits at the same angular velocity as the resonant potential in the expansion of the perturbing potential due to Charon, the so-called pattern speed. In resonances of this type there are forced eccentricities for Nix and Hydra whose magnitudes vanish with e_c , but these do not lead to crossing orbits or other unstable configurations.

The initial orbit of Charon typically has an eccentricity $e_c \sim 0.5$ in the simulations by Canup (2005), and a large eccentricity must be maintained throughout most of the migration of Charon in order

to maintain the stability of the co-rotation resonance and leave the satellites near these resonances after e_c damps. The ratio of the effect of tidal dissipation in Pluto trying to increase e_c to that in Charon trying to reduce e_c in eqn [58] in Appendix B is given by

$$\left| \frac{(\mathrm{d}e_c/\mathrm{d}t)_P}{(\mathrm{d}e_c/\mathrm{d}t)_C} \right| = \frac{19 R_P}{28 R_C} \frac{\mu_C Q_C}{\mu_P Q_P} = 1.30 \frac{\mu_C Q_C}{\mu_P Q_P} \quad [10]$$

where the Love number for a homogeneous sphere, $k_2 = 1.5/(1 + 19\mu/\rho g R) \approx 3\rho g R / 19\mu$ has been used, with μ being the rigidity and g the surface gravity, along with 1164 and 604 km for the radii of Pluto and Charon. If the ratio in eqn [10] exceeds unity, Charon's eccentricity will tend to grow rather than damp. Recall that in the most probable scenario in the Canup (2005) simulations, where Charon's internal temperature is raised only slightly from the impact, Pluto is likely to have differentiated if it was not already so. A soft Pluto compared to a more rigid Charon means it is most likely that $\mu_C Q_C \gg \mu_P Q_P$, at least initially, since both μ_P and Q_P are likely to be smaller in a softer Pluto. We conclude that the existence of the additional satellites in Charon's orbital plane can be accounted for within the paradigm of the large, oblique collision that created Charon. It would be prudent however to perform a numerical simulation to ensure the tidal evolution of the system that is inferred. We should also note that our use of eqn [58] from Appendix 2 neglects higher order terms, which are important for the high eccentricity implied for Charon's orbit. The inclusion of these terms are not expected to change our conclusions qualitatively.

10.14.9 Irregular Satellites

The irregular satellites that orbit all of the major planets occupy large orbits with significant inclinations and eccentricities. Recent observations have increased the number of known irregular satellites to 96, (Gladman *et al.*, 1998, 2000, 2001; Holman *et al.*, 2004). They often extend to distances from their primaries that are close to the stability limits of about half of the planet's Hill sphere radius (Hamilton and Burns, 1991, Hamilton and Krivov, 1997). Similar orbital elements for members of clusters of the irregular satellites imply breakup events after the primary body was captured (e.g., Gladman *et al.*, 2001). The only viable means of establishing these systems of large orbits with high eccentricities and inclinations is through capture from the large

number of planetesimals orbiting the Sun early in solar system history (e.g., Burns, 1986). Capture can result from a loss of energy by the planetesimal while it is within the planet's Hill sphere. Such energy loss can be either from gas drag in an extended planetary atmosphere (Pollock *et al.*, 1979), from a collision between two unbound planetesimals leaving one with sufficient energy loss to be bound, or from a collision with a smaller object already in stable orbit about the primary (Colombo and Franklin, 1971). The gas drag mechanism for at least the progenitor of the Himalia group of prograde irregular satellites orbiting Jupiter has been explored in detail by Cuk and Burns (2004), where it is shown that such capture must occur near the end of a waning gas disk lifetime to avoid losing the satellites to the planet. Gas drag no doubt resulted in the capture of many planetesimals into stable orbits about the major planets, but only the very last ones have survived. The occupancy of orbital resonances by two of the retrograde satellites of Jupiter implies that at least a thin disk was necessarily in place at the time of the capture to allow a weak gas drag to bring the satellites to the resonances (Saha and Tremaine, 1993). An alternative mechanism of capture involves a rapid growth of the primary, while a planetesimal is temporarily trapped in the Hill sphere, to expand the region of stable orbits (Heppenheimer and Porco, 1977; Saha and Tremaine, 1993). The binary exchange capture for the irregular satellites that is highly probable for the capture of Triton (Agnor and Hamilton, 2004, 2006) is probably not as effective for the capture of the distant irregular satellites, since an irregular satellite is less likely to have ever been close enough to the planet for the binary Hill sphere to have been smaller than the separation of the binary members.

Capture of satellites should lead to a random distribution of orbital inclinations relative to the planet's orbital plane (\sim ecliptic), but there are few irregulars known with inclinations between 47° and 141° . Carruba *et al.* (2002) give a lucid explanation of the scarcity of satellites with orbital inclinations within this range. Solar perturbations dominate the variations in the orbits of these satellites, and in an analytic secular theory of the restricted three body problem involving perturbations of the satellite orbit by the lowest order solar tidal term, the motion is averaged over the mean motions of both the planet and the satellites. In this approximation, the component of the satellite's angular momentum that is perpendicular to the planetary orbit plane, $\sqrt{Gm_p a_s(1 - e^2)} \cos i$, is conserved, so that the variation in e and i are linked.

When the satellite orbit plane is closest to the planetary orbit plane, the eccentricity is maximal. The variations in e and i are thus of opposite algebraic sign for prograde orbits ($i < 90^\circ$), but for retrograde orbits ($i > 90^\circ$) increasing i brings the orbit closer to the planetary orbit plane so variations in e and i have the same algebraic sign. For $i < i_{\text{crit}} = 39.2^\circ$ or $i > i_{\text{crit}} = 140.8^\circ$, the amplitude of the variations in e and i are limited, but for inclinations between the values of i_{crit} , large variations in e can drive the periapse into the region of the regular satellites, or the apoapse close to R_H when the argument of periapse ω is circulating. Either consequence can lead to the elimination of the satellite. The only way a high-inclination satellite is likely to survive is if ω is librating about 90 or 270° (Kozai resonance) where variations in the eccentricity are more modest. By estimating the fraction of the available phase space for a captured satellite to be in Kozai resonance Carruba *et al.* predict that perhaps 10% of captured satellites should be in stable Kozai resonances at high inclination. The irregular satellites S/2003 J10, Kiviuq (S/2000 S5), and Ijiraq (S/2000 S6) are apparently currently librating in Kozai resonances. But at least Kiviuq is close to the chaotic zone separating libration from circulation, so its libration may be only temporary (Carruba *et al.*, 2003, 2004).

Numerical integrations of the complete equations of motion of irregular satellites, including the perturbations of the other major planets, show that highly inclined orbits outside the Kozai resonance are more unstable for orbit planes further from the planetary orbit plane between the values of i_{crit} (Carruba *et al.*, 2002). As the extremes of the variable inclinations approach the limits of the zone of exclusion, the satellites can survive for longer periods of time, but the unstable region seems to approach the observed limits asymptotically as the integrations progress. The nonrandom distributions of orbital inclinations and values of the other observed orbital parameters of the irregular satellites are therefore completely consistent with the satellites having been captured into random orbits from the solar nebula population of planetesimals by any of the means discussed above.

Acknowledgments

Special thanks are due to Robin Canup who kindly provided me with data files for some of the figures and to Robin and Bill McKinnon for providing thorough and thoughtful reviews of the manuscript. Support by the NASA Origins of Solar Systems

Program under grant NAG5 13149 and Planetary Geology and Geophysics Program under grant NNG05GK58G is gratefully acknowledged.

Appendix 1: Accretion Disks

The regular satellites of the major planets are in nearly circular orbits in the planes of the planets' equators. This means these satellites had to have formed in dissipative disks of gas and particles that settle to the mid-plane of the disk before the accretion of the satellites occurs. Given that these satellites form in such a disk during the formation of the planet, it is important to understand the nature of the accretion in such disks. Our ultimate goal is the determination of the timescales for accretion of such satellites as a function of the disk and particle parameters, for these timescales constrain the current observable properties of the regular satellites. Once formed, the satellites interact with the disks primarily through gas drag when they are small, and through density waves generated in the adjacent regions of the disk at mean motion commensurabilities with the satellite when they are large. If the mid-plane pressure of the disk decreases with distance from the primary, gas drag causes a satellite to spiral toward the planet. Generally the same type of disk causes a similar spiral toward the planet through interaction of the satellite with density waves generated in the disk. However, if the surface mass density gradient of the disk is sufficiently steep in the radial direction, the density waves interior to the satellite dominate over those exterior and the satellite will spiral away from the planet.

Accretion Details

First we consider the idealized case of an isolated spherical body m_1 of radius R_1 in a swarm of particles with a distribution $n(m_2, v)dm_2 dv$ representing the number density of particles of mass within dm_2 of m_2 and speeds within dv of v . If the relative velocities are isotropic, $n(m_2, v) \sin \theta d\theta d\phi dm_2 dv / 4\pi$ represents the fraction of these particles with velocities aimed within $d\theta d\phi$ of spherical polar coordinates θ, ϕ in a coordinate system where the mean particle velocity is zero. One of these particles of radius R_2 and speed v far from m_1 will collide with m_1 during an encounter if the impact parameter is $< R_g$, where

$$R_g^2 = (R_1 + R_2)^2 \left(1 + \frac{v_e^2}{v^2}\right) \quad [11]$$

with $v_e = \sqrt{2G(m_1 + m_2)/(R_1 + R_2)}$, the escape velocity of m_2 from m_1 when they are in contact. Equation [11] is found by conserving energy and angular momentum when m_2 just grazes m_1 during the encounter. The factor $(1 + v_e^2/v^2) = F_g$ is called the gravitational enhancement factor for the collisional cross section of m_1 for particles of radius R_2 and speed v relative to m_1 . The mass flux of particles with mass within dm_2 of m_2 and velocity within dv of v directed within $d\theta d\phi$ of θ, ϕ onto m_1 is $\pi R_g^2 n(m_2, v) m_2 v dm_2 dv \sin \theta d\theta d\phi / 4\pi$, such that

$$\dot{m}_1 = \int_{m_2} \int_v \pi n(m_2, v) m_2 v (R_1 + R_2)^2 \times \left(1 + \frac{2G(m_1 + m_2)}{(R_1 + R_2)v^2}\right) dm_2 dv \quad [12]$$

is the rate of growth of m_1 , if we assume each m_2 that collides is accreted. In eqn [12], we have assumed the velocities are isotropic relative to m_1 and integrated over the solid angle, and use $R_2 = (3m_2/4\pi\rho_p)^{1/3}$, with ρ_p being the mass density of the individual particles. Because $n(m_2, v)$ is not very well constrained, eqn [12] is often replaced by the much simpler form,

$$\dot{m}_1 \approx \frac{\sigma_s}{2H_s} v \pi (R_1 + R_2)^2 \left(1 + \frac{v_e^2}{v^2}\right) \quad [13]$$

where R_2 , v_e , and v are now suitably averaged values. We have replaced the spatial mass density of the particles by $\sigma_s/2H_s$, where σ_s is the surface mass density of solids and $2H_s$ is the disk thickness. H_s depends on the velocity dispersion v , and can be estimated by

$$0.5v_\perp^2 = \int_0^{H_s} (GMz/r^3) dz = 0.5\Omega_K^2 H_s^2 \quad [14]$$

where $v_\perp = Kv$ ($K = O(1)$) is the perpendicular component of the averaged velocity, G is the gravitational constant, M is the mass of the primary, r is the distance from the primary, z is the vertical distance from the mid-plane, and $\Omega_K = \sqrt{GM/r^3}$ is the Kepler orbital angular velocity of a test particle at distance r from the primary. GMz/r^3 is just the local vertical component of the gravitational acceleration due to the primary. Hence, $H_s = v_\perp/\Omega_K = Kv/\Omega_K$. This value of H_s is generally smaller than the gas pressure scale height c_s/Ω_K (derived below), with c_s

being the speed of sound, because of settling of particles toward the mid-plane.

From eqn [13], we can write

$$\dot{R}_1 = \frac{\sigma_s \Omega_K}{8K \rho_p} \left(1 + \frac{R_2}{R_1} \right)^2 (1 + 2\theta_s) \quad [15]$$

where ρ_p is the mass density of the particles as distinguished from ρ_s , the spatial density of solids in the disk, and $\theta_s = v_e^2/2v^2$ is the Safronov number ($F_g = 1 + 2\theta_s$). If we assume $R_2 \ll R_1$ and ignore factors of order unity, we can write the timescale for the growth of a body as (Canup and Ward, 2002)

$$\tau_A = \frac{R_1}{\dot{R}_1} \approx \frac{\rho_p R_1}{\Omega_K \sigma_s F_g} \sim 50 \text{ yr} \left(\frac{R_1}{2500 \text{ km}} \right) \left(\frac{\rho_p}{2 \text{ g cm}^{-3}} \right) \times \left(\frac{10}{F_g} \right) \left(\frac{3 \times 10^3 \text{ g cm}^{-2}}{\sigma_s} \right) \left(\frac{r}{15 R_J} \right)^{3/2} \quad [16]$$

where the numerical form is for the accretion of Ganymede in a minimum mass disk of particles. The surface mass density σ_s is estimated by spreading Ganymede's mass over an annulus extending from the midpoint between Ganymede's and Europa's orbits to the midpoint between Ganymede's and Callisto's orbit to yield $\sigma_s = 3300 \text{ g cm}^{-2}$. With $\rho_p \approx 2 \text{ g cm}^{-3}$ and an orbital radius of $15 R_J$ we arrive at $\tau_A \sim 50$ years when $F_g = 10$. This is a lower bound on the timescale for growth of Ganymede, since σ_s is reduced as the growth proceeds, but it is indicative of the very short timescales for accretion in minimum mass disks.

Safronov (1969), by transforming from relative to random velocities and assuming the vertical and horizontal velocity dispersions are equal, finds $K = \pi/4\sqrt{2}$. Safronov also assumed that the velocity dispersion of the accreting particles remained comparable to the escape velocity of the largest embryos as R_1 increased. In this case $\theta_s = O(1)$, and R_1 increases linearly with time if the surface mass density σ_s is maintained in the vicinity of m_1 . This linear increase in the radius of an accreting body is called orderly growth. In this case $F_g \approx 3$, which is not that much smaller than our choice of $F_g = 10$ in eqn [16]. Runaway growth can make these accretion timescales even shorter.

The assumption by Safronov that the velocity dispersion of the smaller particles was always near the escape velocity of the largest bodies proved to be incorrect in numerical simulations by Wetherill and Stewart (1989). As long as most of the mass of the disk was in the smaller bodies, the velocity dispersion was nearly the escape velocity of the smaller particles. Increases in the velocities of small particles by close

encounters with the larger bodies were damped by interactions with members of the swarm. The Safronov number is thereby increased by orders of magnitude from that assumed in orderly growth and led to what is called runaway growth. For v fixed near the escape velocity of the smaller bodies, the rate of growth in eqn [15] becomes proportional to v_e^2 which, for constant density, increases as R_1^2 . The change in scaling results in the largest planetesimal growing much faster than any other, and it 'runs away'. For the two largest planetesimals of the swarm of radii $R_1 > R_2$,

$$\frac{d}{dt} \left(\frac{R_1}{R_2} \right) = \frac{R_1}{R_2} \left(\frac{\dot{R}_1}{R_1} - \frac{\dot{R}_2}{R_2} \right) \approx \frac{R_1}{R_2} (R_1 - R_2) \quad [17]$$

shows that if initially $R_1 > R_2$ the ratio R_1/R_2 grows as long as the mass available for accretion has not decreased substantially.

The timescale for accretion in eqn [16] is reduced by the larger value of F_g . Both orderly growth and runaway growth timescales depend on the maintenance of the surface mass density of the disk near the accreting satellite, and that most of the disk mass is in the small particles. These timescales increase as the nearby material is exhausted, and must be eventually governed by how rapidly particles can be brought into the so-called feeding zone of the growing satellite. Still, numerical simulations of lunar accretion in a particle disk verify the rapid accretion timescales even while accounting for the growth of the average particle size during the process. (Ida *et al.*, 1997; Kokubo *et al.*, 2000). Mosqueira and Estrada (2003a, 2003b) determine accretion timescales governed by how rapidly gas drag can bring larger embryos to the feeding zone of the satellite.

Although some stages of the accretion process in disks remain unclear, once small solid objects have formed or are deposited in a flat dissipative medium, their accumulation into a coplanar system of satellites seems reasonably well understood in spite of continued haggling over details. Timing of the formation depends on the temperature history of the disk when ices are the predominant constituents of the forming satellites. Once formed, the satellites will spiral toward the primary because of gas drag and interaction with density waves in the gaseous part (Type 1 drift) (Ward, 1997). It is thus likely that at least the large regular satellites formed during the last stages of accretion of their host planet as the gas in the disk and the solar nebula was being dissipated. The small equatorial satellites residing close to their primaries

have most likely been repeatedly disintegrated and reformed in a gas-free environment by planetesimals remaining in the heliocentric swarm. Instabilities in the distribution of large objects lead to mergers and eventual separation of the remaining objects by more than several Hill sphere radii, where the Hill sphere radius (sphere of influence) is defined as the distance from the satellite to its inner Lagrange point ($R_H \approx r(M_s/3M_p)^{1/3}$).

Refinements to the accretion process in disks take account of the effect of the primary and depletion of accreting particles in the feeding zone of a growing object (Wetherill and Cox, 1984, 1985; Lissauer, 1987; Greenzweig and Lissauer, 1990, 1991), where an upper bound on the gravitational enhancement factor of about 1000 is determined. Gas drag replenishes the feeding zone and thereby prolongs the accretion of a body in a disk, but only a fraction of such particles (10–40%) brought into the feeding zone by gas drag are actually accreted (Kary *et al.*, 1993; Kary and Lissauer, 1995).

Viscous Spreading

An inherent property of disks orbiting a primary is that they tend to spread as angular momentum is exchanged between adjacent regions because of the Kepler shear. Gravitational, magnetic, and viscous torques and wave transport with shock dissipation are all possible candidates for angular momentum transport (Balbus, 2003). Viscous interchange results from the Kepler shear in the motion of the material orbiting the primary. If we represent the shear stress as due to a viscosity η , we can write the shear stress at a distance r from the primary $w_{r\phi} = \eta|dv_\phi/dr|_{\text{rot}} = \eta r|d\Omega_K/dr| = 3\eta\Omega_K/2$, where $|dv_\phi/dr|_{\text{rot}}$ is the radial gradient of the azimuthal velocity in the frame rotating at the Kepler angular velocity at distance r . For a gas,

$$\eta = \frac{m\bar{v}}{3A} \approx \frac{mc_s}{3A} \quad [18]$$

where m is the mean molecular mass, $\bar{v} = \sqrt{8kT/\pi m} \approx c_s$ is the mean molecular speed with k being Boltzmann's constant, T is the temperature, A is the collision cross-section of the molecules, and c_s is the sound speed (e.g., Sears and Salinger, 1975). Most of the regular satellites around the major planets have a substantial fraction of ice. So if ice is to be in solid form as satellites, disk temperatures must be less than ~ 200 K (depending on the pressure), so for molecular

hydrogen, $c_s \approx 10^5$ cm s $^{-1}$. The kinematic viscosity $\nu = \eta/\rho_g = 2H_g\eta/\sigma_g \approx c_s H_g m/(A\sigma_g)$, where ρ_g is the gas density, σ_g is the surface mass density of the gas, and H_g is the half-thickness of the gas disk.

For the gaseous disk we choose H_g to be the scale height of the disk which is determined by isostacy in the direction perpendicular to the disk, $dP/dz = -\rho_g g_z = -mPg_z/kT = -mPz\Omega_K^2/kT$, where P is the pressure, $g_z = GMz/r^3 = z\Omega_K^2$ is the component of gravity in the z direction. Solution of this equation gives $P = P_0 \exp -z^2/H_g^2$, where $H_g = \sqrt{2kT/m}/\Omega_K \approx c_s/\Omega_K$ is the isothermal scale height. As an example we can estimate the timescale for viscous transfer of angular momentum in a disk surrounding Jupiter that has the minimum mass for accreting Ganymede at Ganymede's distance from Jupiter augmented with volatiles to reach solar composition. The amount of solids in the disk necessary to make Ganymede corresponds to a surface density of 3300 g cm $^{-2}$ spread from the mid-point between Ganymede and Europa and the midpoint between Ganymede and Callisto is thereby increased to $\sim 10^5$ g cm $^{-2}$ such that the kinematic viscosity $\nu \approx 3 \times 10^5$ cm 2 s $^{-1}$. The timescale for viscous transfer of angular momentum at Ganymede's distance from Jupiter ($\sim 10^{11}$ cm) is $r^2/\nu = 1.2 \times 10^9$ years. This time is so long compared to other timescales associated with a circumplanetary disk that molecular viscosity cannot be important in the disk evolution.

If the gas is turbulent, a turbulent viscosity is invoked to increase the efficiency of angular momentum transport. In this approximation, the molecular mass in eqn [18] is replaced by the mass of a typical turbulent eddy $\sim \rho_g l^3$, and the cross-section A by $\sim l^2$, where l is a typical eddy size. Finally \bar{v} is replaced by the mean random velocity of the turbulent eddy v_t , so $\eta_t = \rho_g l v_t$. Next $v_t = \alpha c_s$, where α is a free parameter, $l \sim H_g = c_s/\Omega_K$, so that the kinematic viscosity appropriate for a turbulent medium is (Shakura and Sunyaev, 1973)

$$\nu_t = \frac{\alpha c_s^2}{\Omega_K} \quad [19]$$

The free parameter α is used to indicate the strength of the nonmolecular viscous coupling, and the models are often called ‘ α disks’. Choices of α are usually based on unsupported assertions about the strength of the turbulence. Lower bounds on the magnitude of angular momentum transport follow from the masses of the giant planets. An equivalent viscosity must be sufficient to prevent the forming planet from

opening a gap in the proto-planetary disk before it approaches the observed mass (Lin and Papaloizou, 1985). Typically assumed values of $\alpha = O(10^{-3})$ in applications. For shear stress $w_{r\phi}$, the torque across a cylindrical surface of radius r centered on the primary is $T_\nu = w_{r\phi} 2\pi r^2 2H_g = 6\pi\eta\Omega_K r^2 H_g = 3\pi\alpha\sigma_g c_s^2 r^2$ for an α disk. For the example of a minimum mass disk at Ganymede's distance from Jupiter with $c_s = 10^5 \text{ cm s}^{-1}$ and $\alpha = 10^{-3}$, $\nu_t \approx 10^{12} \text{ cm}^2 \text{s}^{-1}$ and the timescale for viscous spreading ~ 400 years.

If we consider the mass dm in the cylindrical region of radial extent dr with specific angular momentum h , we have $D(b dm)/Dt = -(dT_\nu/dr)dr$ to balance the net torque on the cylinder with the change in the angular momentum, where $D/Dt \equiv \partial/\partial t + \mathbf{v} \cdot \nabla$ is the total derivative. In a steady-state disk,

$$v_r \frac{dm db}{dr dr} = F_r \frac{db}{dr} = - \frac{dT_\nu}{dr} \quad [20]$$

where F_r is the radial mass flux. For the general case where $T_\nu = 0$ at $r = 0$ and as $r \rightarrow \infty$, and with $db/dr > 0$ as for Kepler motion, we see that dT_ν/dr must be positive at small r and negative at large r so that mass moves inward for small r and outward for large r . Angular momentum always moves outward.

Particle disks also spread from angular momentum transfer via particle collisions. For a disk of particles of mass m_p and velocity dispersion \bar{v} about the Kepler velocity at distance r from the primary, the equivalent kinematic viscosity is given by $\nu_s = \eta_s/\rho_s = m_p \bar{v}/(3A\rho_s) = 2m_p \bar{v} H_s / (3A\sigma_s) = 4H_s \bar{v} \rho_p R_p / (9\sigma_s)$, where σ_s is the surface mass density of particles in the disk, R_p is the particle radius, and ρ_p is the particle density (distinguished from the spatial mass density of particles ρ_s). For a disk of 100 m particles corresponding to the minimum mass particle disk at Ganymede's distance from Jupiter, and assumption $\bar{v} = v_{\text{esc}}$ of two touching particles, $H_s = K v_{\text{esc}}/\Omega_K$ and $\rho_p = 1 \text{ g cm}^{-3}$, ν_s is only a little over $80 \text{ cm}^2 \text{s}^{-1}$. So low dispersion particle disks spread very slowly with timescale $r^2/\nu = 4.5 \times 10^{12}$ years. But inside the Roche radius in the lunar case, gravitational instabilities cause patches of higher-density material that are repeatedly sheared apart, thereby yielding enhanced collisional dissipation with an effective viscosity of $\nu \sim (\pi G \sigma_s)^2 / \Omega_K^3$ and a resulting timescale $r^2/\nu_t \sim 1$ year (Ward and Cameron, 1978). This rapid spreading has been confirmed by N-body simulations (Takeda and Ida, 2001). But self-heating during this short timescale spreading is so intense, at least for the proto-lunar

disk, that more silicates are vaporized thereby decreasing the effective viscosity and allowing longer timescales for spreading (Thompson and Stevenson, 1988). So the long timescales for spreading of particulate disks from the simple molecular viscosity analogy are not likely to prevail.

Gas Drag

A forming satellite assumes a Kepler orbit about the primary, but the gas in the disk has its central acceleration reduced by a radial pressure gradient, and thereby orbits slower than the satellites. The satellite thus experiences a continuous head wind that causes it to spiral toward the primary (Whipple, 1964; Weidenschilling, 1977). The gas motion is governed by the equation

$$\rho_g \frac{d\mathbf{v}}{dt} = -\rho_g \nabla \Phi - \nabla P \quad [21]$$

where Φ is the gravitational potential per unit mass and P is the pressure. In a steady state, cylindrical coordinates $\vec{r}, \vec{\phi}, \vec{z}, \vec{r}, \vec{z}$ are all zero, and the r component of eqn [21] is

$$-\rho_g r \dot{\phi}^2 = -\rho_g \frac{\partial \Phi}{\partial r} - \frac{\partial P}{\partial r} \quad [22]$$

Now $P = \rho_g kT/m = \sigma_g c_s^2 / 2H_g = \sigma_g c_s \Omega_K / 2$ (m = mean molecular mass). Often the surface mass density and mid-plane temperatures are represented by power laws (e.g., Weidenschilling, 1977) like $\sigma_g = \sigma_{g0}(r_0/r)^p$ and $T = T_0(r_0/r)^q$, where σ_{g0} and T_0 are values at distance r_0 from the primary. Then $\partial P/\partial r = -K\sigma_g \Omega_K c_s / r$, where $K = O(1)$ depends on the choices for p and q . With $\Phi = -GM_p/r$, solution of eqn [22] yields the gas angular velocity,

$$\dot{\phi}^2 \equiv \Omega^2 = \frac{GM_p}{r^3} - \frac{K\sigma_g \Omega_K c_s}{\rho_g r^2} = \Omega_K^2 - \frac{2Kc_s^2}{r^2} \quad [23]$$

from which

$$\Omega_K - \Omega \approx \frac{Kc_s^2}{r^2 \Omega_K} \quad [24]$$

where $\Omega_K + \Omega \approx 2\Omega_K$ has been used.

For circularly orbiting material the gas drag force is given by

$$F_d = -C_d \pi R^2 \rho_g v_{\text{rel}}^2 \quad [25]$$

where C_d is the drag coefficient, which depends on the Reynolds number $R = 2v_{\text{rel}} R \rho_g / \eta$, where η is the molecular viscosity, and $v_{\text{rel}} = r(\Omega_K - \Omega) = Kc_s^2 / r \Omega_K$ is the relative velocity between the forming

satellite and the gas, that is, the head wind. The Reynolds number dependence of C_d (e.g., Prandtl, 1952) leads to $C_d = 20/R$ (Stokes drag) for $R \lesssim 0.5$ and $C_d \approx 0.43$ for $R \gtrsim 500$. (See Peale (1993) for an application of gas drag with full Reynolds number dependence for C_d .) The torque on the satellite is rF_d , which when equated to the time derivative of the angular momentum given by $m_p\sqrt{GM_p r}$ yields an expression for \dot{r} . The timescale for gas drag changes in the orbit is then

$$\begin{aligned} \tau_{GD} \equiv \frac{r}{\dot{r}} &= \frac{4}{3K^2} \frac{1}{C_d} \frac{\rho_p R}{\Omega_K \sigma_g} \left(\frac{r \Omega_K}{c_s} \right)^3 \\ &\sim 10^3 \text{ yr} \left(\frac{10}{C_d} \right) \left(\frac{R}{2500 \text{ km}} \right) \left(\frac{\rho_p}{2 \text{ g cm}^{-3}} \right) \\ &\times \left(\frac{0.1}{c_s / r \Omega_K} \right)^3 \left(\frac{3 \times 10^5 \text{ g cm}^{-2}}{\sigma_g} \right) \left(\frac{r}{15R_J} \right)^{3/2} \end{aligned} \quad [26]$$

where we have used $K = 3/4$ for $p = q = 1$, and the numerical form is for Ganymede (Canup and Ward, 2002), where the minimum mass particle disk has been augmented by ~ 30 times as much gas to yield solar composition.

Migration from Density Waves

Satellites also interact with disks by creating spiral density waves at the commensurabilities of the mean motions in the disk with that of the satellite (e.g., Ward, 1997). The resulting distribution of mass in the density waves exert gravitational torques on the satellite that tend to push the satellite away from the site of the density wave. Hence, the density waves generated inside the satellite orbit tend to increase its angular momentum, whereas those outside tend to decrease it. Simultaneously, the ring material receives or gives up angular momentum, which tends to push the ring material away from the satellite. The distribution of the resonances outside the satellite versus those inside lead to the outside density waves dominating, such that the net effect is for the satellite to spiral toward its primary, called Type I drift, with a timescale given by (Ward, 1997; Canup and Ward, 2002)

$$\begin{aligned} \tau_I \equiv \frac{r}{\dot{r}} &\approx \frac{1}{C_a \Omega_K} \left(\frac{M_p}{M_s} \right) \left(\frac{M_p}{r^2 \sigma_g} \right) \left(\frac{c_s}{r \Omega_K} \right)^2 \\ &\sim 10^2 \text{ yr} \left(\frac{3}{C_a} \right) \left(\frac{2500 \text{ km}}{R} \right)^3 \left(\frac{2 \text{ g cm}^{-3}}{\rho_p} \right) \\ &\times \left(\frac{C_s / r \Omega_K}{0.1} \right)^2 \left(\frac{3 \times 10^5 \text{ g cm}^{-2}}{\sigma_g} \right) \left(\frac{15R_J}{r} \right)^{1/2} \end{aligned} \quad [27]$$

where C_a is a torque asymmetry parameter that is a function of the disk's radial surface density and temperature profiles (e.g., Ward, 1997; Tanaka *et al.*, 2002), M_p and M_s are the planet and satellite masses, respectively, and r is the distance from the M_p center of mass. The last term in eqn [27] is $(H_g/r)^2$. The numerical representation is that centered on Ganymede in a minimum-mass Jovian disk.

The repulsion of the gas disk by the satellite can eventually open a gap in the disk if (e.g., Ward and Hahn, 2000)

$$\frac{M_s}{M_p} > c_\nu \sqrt{\alpha} \left(\frac{c_s}{r \Omega_K} \right)^{5/2} \sim 10^{-4} c_\nu \left(\frac{\alpha}{10^{-3}} \right)^{1/2} \left(\frac{c_s / r \Omega_K}{0.1} \right)^{5/2} \quad [28]$$

where $c_\nu = O(1 - 10)$. This form of the criterion comes from Ward and Canup, where they point out that the limiting mass is just above Ganymede's mass for $c_\nu = 1$ and $\alpha = 10^{-3}$. If a gap is opened, the Type I drift is replaced by Type II, whose timescale is now the disk viscous timescale (Ward, 1997)

$$\begin{aligned} \tau_{II} &= \frac{r^2}{\nu_t} = \frac{(r/H_g)^2}{\alpha \Omega_K} \\ &\sim 300 \text{ yr} \left(\frac{10^{-3}}{\alpha} \right) \left(\frac{0.1}{H_g/r} \right)^2 \left(\frac{r}{15R_J} \right)^{3/2} \left(\frac{M_J}{M_p} \right)^{1/2} \end{aligned} \quad [29]$$

where again the numerical representation is centered on Ganymede and M_p is normalized by the Jupiter mass M_J . Whether or not a gap is opened by a satellite depends on the value of α and the scale height of the disk at the satellite location.

Thermodynamics

The temperature distribution in the disk is of fundamental importance in determining nearly all of the parameters associated with accretion within the disk and whether or not the accreted particles survive migration toward the primary. The growing satellites will be near the mid-plane of the disk, so the mid-plane temperature T_e must be known. This in turn is determined by an energy balance equation in steady-state disks, where sources of energy include the gravitational energy deposited by the flux of incoming material, the viscous dissipation within the disk, radiation from the still warm primary, and radiation from the surrounding solar nebula. The sink is radiation from upper and lower surfaces of the disk at some effective temperature, T_e . If the accretion of the

satellites is assumed to take place in a disk with no accretion from the solar nebula, the gravitational source of energy now comes from the redistribution of mass during viscous evolution, and the internal energy of the disk is reduced as it cools.

For optically thick disks $T_e^4 = 2T_c^4/(1 + 3\tau/2) \approx 4T_c^4/3\tau$, whereas for optically thin disks $T_e^4 \approx 4\tau T_c^4$ (e.g., Shapiro and Teukolsky, 1983), where $\tau = \int_0^{H_g} \kappa \rho_g dz \approx \bar{\kappa} \sigma_g / 2$ is the optical depth from mid-plane to the surface, with κ being the Rosalind mean opacity ($\text{cm}^2 \text{g}^{-1}$) (e.g., Schwarzschild, 1958). An example of a surface mass distribution in a steady-state disk from Canup and Ward (2002) (Lynden-Bell and Pringle, 1974) is

$$\begin{aligned} \sigma_g(r) &\approx \frac{4\dot{M}}{15\pi\nu_t} \left[\frac{5}{4} - \sqrt{\frac{r_c}{r_d}} - \frac{1}{4} \left(\frac{r}{r_c} \right)^2 \right] \quad \text{for } r < r_c \\ &\approx \frac{4\dot{M}}{15\pi\nu_t} \left[\sqrt{\frac{r_c}{r}} - \sqrt{\frac{r_c}{r_d}} \right] \quad \text{for } r \geq r_c \end{aligned} \quad [30]$$

where the kinematic viscosity $\nu_t = \alpha c_s^2 / \Omega_K$ is the approximation of Shakura and Sunyaev (1973) for a turbulent medium discussed above, r_c is the radius inside of which the mass flux \dot{M} is deposited uniformly and r_d is the edge of the disk. Another example is a disk with constant mass flux toward the primary throughout the region where the satellites are forming (e.g., Lynden-Bell and Pringle, 1974),

$$\sigma_g(r) = \frac{\dot{M}}{3\pi\nu_t} \left[1 - \sqrt{\frac{R_p}{r}} \right] \quad [31]$$

if the viscous couple vanishes at the primary's surface of radius R_p .

Figure 4 shows the Rosalind mean opacities appropriate for low temperatures in an accretion disk. The opacity due to Bell and Lin (1994) accounts for ice grains at low temperatures and ‘metallic’ grains at high temperatures. Ice particles dominate the opacity up to temperatures slightly above 160 K with approximately a T^2 power law dependence. Above this temperature the ice evaporates with a sharp drop in opacity and ‘metallic’ grains dominate the opacity for larger T . The Pollack *et al.* (1994) opacities include a distribution of organic particles that maintain a high opacity in the region beyond the ice point. These opacities can be modeled by writing $\kappa = \kappa_0 T^\xi$, but different values of the exponent ξ apply in different parts of the disk. There have been no studies of the formation of icy satellites with this accounting of the variations in the disk opacities. An example of the energy balance

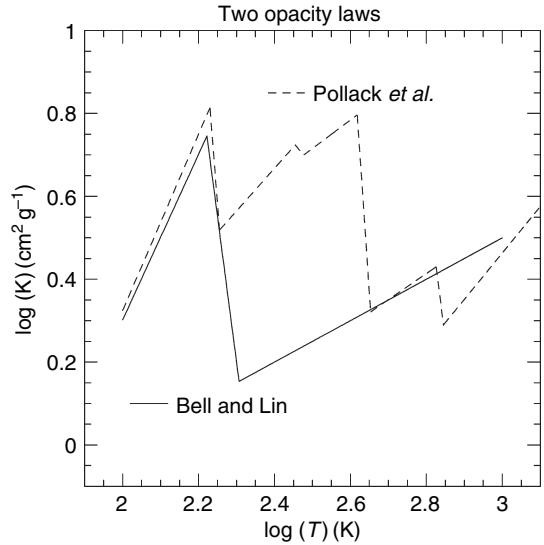


Figure 4 Power law approximations to the Rosseland mean opacities for a cold accretion disk (Bell and Lin, 1994; Pollack *et al.*, 1994).

equation for a steady-state disk at distance r from the primary is

$$\begin{aligned} \sigma_{SB} T_p^4 &\left[\frac{2}{3\pi} \left(\frac{R_p}{r} \right)^3 + \frac{1}{2} \left(\frac{R_p}{r} \right)^2 H \left(\frac{d \ln H}{d \ln r} - 1 \right) \right] \\ &+ F_{\text{grav}}(r) + F_{\text{vis}}(r) \\ &= 2\sigma_{SB} \left(T_e^4 - \frac{\beta}{4\pi} T_{\text{neb}}^4 \right) \end{aligned} \quad [32]$$

where the first bracketed term on the left is the radiation deposited per unit area from the warm planet (Ruden and Pollock, 1991), $F_{\text{grav}}(r)$ is the energy deposited due to the accretion, and $F_{\text{vis}}(r)$ is the energy deposited per unit surface area from viscous interactions. The fraction of solid angle subtended by the solar nebula is $\beta/4\pi$. The most extreme model dependences are in the terms F_{grav} and F_{vis} . There are no constraints on F_{grav} from simulations, and F_{vis} depends on the unknown value of α . The importance of the opacity κ for the midplane properties cannot be overemphasized.

Appendix 2: Tides

Tidal dissipation can drastically change the orbital configuration of the regular natural satellites from the time of their formation. So it is imperative to understand this evolution, which puts constraints on the conditions of satellite formation. We therefore outline a theory of tidal effects on satellite orbits (and rotations) in this appendix.

The potential of mass element dm_0 at \mathbf{r} (within m_0) due to a mass M located at \mathbf{R} relative to a coordinate system fixed in the body m_0 is given by

$$\begin{aligned} dV &= \frac{GM dm_0}{r} \\ &= -\frac{GM dm_0}{R} \left[1 + \frac{r}{R} P_1(\cos S) + \left(\frac{r}{R}\right)^2 P_2(\cos S) \right. \\ &\quad \left. + \left(\frac{r}{R}\right)^3 P_3(\cos S) + \dots \right] \end{aligned} \quad [33]$$

where G is the gravitational constant, P_i are the Legendre polynomials, and $\cos S = \mathbf{r} \cdot \mathbf{R}/rR$. The first two terms in the expansion combine to yield an acceleration parallel to \mathbf{R} , so the lowest order tide raising potential per unit mass in m_0 due to M is given by the second harmonic term

$$W_2 = -\frac{GM r^2}{2R^3} [3 \cos^2 S - 1] \quad [34]$$

Typically $r/R \ll 1$ allows neglect of higher-order terms, but if a satellite is close to a primary (e.g., Phobos), higher-order terms contribute significantly to the total tidal interaction (Bills *et al.*, 2005).

For a homogeneous sphere, Love (1944) shows that for a small distortion, the displacement of the surface due to the disturbing potential W_2 is a small factor ϵ_2 times a surface harmonic S_2 of the same degree.

$$\epsilon_2 S_2 = \frac{5}{2g} \frac{W_2(r = a_e)}{[1 + 19\mu/(2\rho g a_e)]} \quad [35]$$

where g is the surface gravity acceleration on m_0 , μ is the rigidity, ρ is the density, and a_e is the radius of the sphere. The potential per unit mass at $r > a_e$ due to this spherical surface harmonic mass distribution is then

$$\begin{aligned} -\frac{3}{5} \left(\frac{a_e}{r}\right)^3 \epsilon_2 S_2 &= -\frac{3}{2[1 + 19\mu/(\rho g a_e)]} \\ &\times \frac{GM a_e^5}{2R^3 r^3} (3 \cos^2 S - 1) \end{aligned} \quad [36]$$

such that if we place a mass m at \mathbf{r} ,

$$V_T = -\frac{k_2 G M m a_e^5}{2R^5 r^5} [3(\mathbf{R} \cdot \mathbf{r})^2 - r^2 R^2] \quad [37]$$

is the potential of m at \mathbf{r} due to a tide raised by M at \mathbf{R} or the potential of M at \mathbf{R} due a tide raised by m at \mathbf{r} . The potential Love number, $k_2 = 1.5/[1 + 19\mu/(2\rho g a_e)]$ refers to the body on which the tide is raised. For centrally condensed bodies, k_2 is reduced. For

example, the fluid Love number for the Earth is 0.934 instead of 1.5 (Lambeck, 1980, p. 29).

Now consider M and m to be the same body. Then $\mathbf{R}(t) = \mathbf{r}(t - \Delta t)$ is the simplest representation of the effects of dissipation of tidal energy in m_0 . That is, \mathbf{R} is the position of m a short time in the past relative to a coordinate system fixed in m_0 , where \mathbf{r} is defined in the same coordinate system. The components of $\mathbf{R} = \mathbf{r}(t - \Delta t)$ are not to be included in differentiations of V_T in determining the tidal forces on m . We distinguish these coordinates with a^* , and write the coordinates of m considered the tide raising body as

$$\mathbf{r}^*(t - \Delta t) = \mathbf{r}^*(t) - \frac{d\mathbf{r}^*}{dt} \Big|_t \Delta t = \mathbf{r}^* + \Delta\mathbf{r}^* \quad [38]$$

If we substitute the right-hand side of eqn [38] into eqn [37] and expand to first order in $\Delta\mathbf{r}^*$, there results

$$V_T = -k_2 G m^2 a_e^5 \left(\frac{1}{r^3 r^{*3}} + \frac{3\mathbf{r} \cdot \Delta\mathbf{r}^*}{r^4 r^{*4}} - \frac{6\mathbf{r}^* \cdot \Delta\mathbf{r}^*}{r^3 r^{*5}} \right) \quad [39]$$

where

$$\begin{aligned} \Delta\mathbf{r}^* &= -\frac{d\mathbf{r}^*}{dt} \Big|_{\text{body}} \Delta t \\ &= -\Delta t \left(\frac{dx^{*'}}{dt} \mathbf{e}'_1 + \frac{dy^{*'}}{dt} \mathbf{e}'_2 + \frac{dz^{*'}}{dt} \mathbf{e}'_3 \right) \end{aligned} \quad [40]$$

where the time derivative is relative to the $x'y'z'$ system of coordinates fixed in the rotating body m_0 . With the generic relation for a vector \mathbf{D} , $(d\mathbf{D}/dt)_{\text{body}} = (d\mathbf{D}/dt)_{\text{space}} - \boldsymbol{\omega} \times \mathbf{D}$, and with $\mathbf{D} \rightarrow \mathbf{r}^*$ and $\boldsymbol{\omega} \rightarrow \dot{\psi}$, where $\dot{\psi}$ is the angular velocity of m_0 , we can write $d\mathbf{r}^*/dt|_{\text{body}} = d\mathbf{r}^*/dt|_{\text{space}} - \dot{\psi} \times \mathbf{r}^*$ and

$$\Delta\mathbf{r}^* \cdot \mathbf{r}^* = -\mathbf{r} \cdot \dot{\mathbf{r}} \Delta t \quad [41]$$

where the derivative on the right-hand side is now relative to inertial space.

As an example we assume m_0 is rotating about the z' axis at angular velocity $\dot{\psi}$ relative to the xyz axes fixed in inertial space with the xy and $x'y'$ planes being coincident, so that $\psi = \psi_0 + \dot{\psi}t$ is the angle between the x and x' axes. Then

$$\Delta\mathbf{r}^* = -(\dot{\psi}y + \dot{x})\Delta t \mathbf{e}_1 + (\dot{\psi}x - \dot{y})\Delta t \mathbf{e}_2 - \dot{z}\Delta t \mathbf{e}_3 \quad [42]$$

The force on m due to the tide it raises on m_0 is found from the negative gradient of V_T in eqn [39] with respect to the unstarred coordinates, after which we can set $r^* = r$ and substitute eqn [41] and [42] into the gradient so obtained to yield

$$\begin{aligned} -\frac{\partial V_T}{\partial x} &= -k_2 G m^2 a_e^5 \left[\frac{3x}{r^8} + \frac{6\mathbf{r} \cdot \dot{\mathbf{r}}x \Delta t}{r^{10}} + \frac{3(\dot{\psi}y + \dot{x})\Delta t}{r^8} \right] \\ -\frac{\partial V_T}{\partial y} &= -k_2 G m^2 a_e^5 \left[\frac{3y}{r^8} + \frac{6\mathbf{r} \cdot \dot{\mathbf{r}}y \Delta t}{r^{10}} + \frac{3(-\dot{\psi}x + \dot{y})\Delta t}{r^8} \right] \\ -\frac{\partial V_T}{\partial z} &= -k_2 G m^2 a_e^5 \left[\frac{3z}{r^8} + \frac{6\mathbf{r} \cdot \dot{\mathbf{r}}z \Delta t}{r^{10}} + \frac{3\dot{x}\Delta t}{r^8} \right] \end{aligned} \quad [43]$$

The gradient $-\nabla V_T$ from eqn [43] is the force on m due to tides raised on m_0 by m . By itself this force will increase m 's orbital semimajor axis and eccentricity and decrease the orbital inclination relative to m_0 's equator plane. Here $k_2 \rightarrow k_2^{m_0}$, the Love number for m_0 , $a_e \rightarrow R_{m_0}$, the equatorial radius of m_0 , $\dot{\psi} \rightarrow \dot{\psi}_{m_0}$, the angular velocity of m_0 , $\mathbf{r} \rightarrow \mathbf{r}_{m_0 m}$, the vector from the center of m_0 to the center of m with corresponding components in inertial space of x, y, z .

The force on m_0 due to tides raised on m by m_0 has the same form as eqn [43] with now $k_2 \rightarrow k_2^m$, the Love number for m , $m \rightarrow m_0$, $a_e \rightarrow R_m$, the radius of m , and $\mathbf{r} = \mathbf{r}_{mm_0} = -\mathbf{r}_{m_0 m}$. To convert this force to the force on m due to tides raised on m by m_0 , the change in sign of the force expressions by Newton's third law is canceled by the change in sign of the coordinates effected by setting $\mathbf{r} = \mathbf{r}_{m_0 m}$ instead of $-\mathbf{r}_{m_0 m}$. So eqn [43] express the force on m due to tides raised on m by m_0 with $k_2 \rightarrow k_2^m$, $m \rightarrow m_0$, $a_e \rightarrow R_m$ and $\mathbf{r} \rightarrow \mathbf{r}_{m_0 m}$.

The torque on m due to tides raised by itself on m_0 is $T_T = \mathbf{r} \times (-\nabla V_T)$, and the torque retarding m_0 's spin is the negative of this, where $\mathbf{r} = \mathbf{r}_{m_0 m}$. Similarly, the torque retarding m 's spin due to tides raised on m by m_0 has the same form, but with $\mathbf{r} = \mathbf{r}_{mm_0}$ and $m \rightarrow m_0$, $a_e \rightarrow R_m$ and $k_2 \rightarrow k_2^m$. This latter torque on m can be represented by

$$\mathbf{T}_T = \frac{3k_2 G m_0^2 R^5}{r^6} (\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_T) (\hat{\mathbf{r}}_T \times \hat{\mathbf{r}}) \quad [44]$$

where we have omitted the sub- and superscripts but remember that the variables apply to m . $\hat{\mathbf{r}}$ is a unit vector toward the tide raising body m_0 , and $\hat{\mathbf{r}}_T$ is a unit vector toward the tidal maximum, which is the sub-primary position on m a short time Δt in the past as above. Eqn [44] is valid for arbitrary inclination of the orbit and equator planes. Similar to the above analysis,

$$\hat{\mathbf{r}}_T = \hat{\mathbf{r}} - \frac{d\hat{\mathbf{r}}}{dt} \Delta t \quad [45]$$

where the time derivative is relative to the body system of coordinates. Replacement of $\hat{\mathbf{r}}_T$ with Eqn [45] yields

$$\mathbf{T}_T = \frac{3k_2 G m_0^2 R^5 \Delta t}{r^6} \hat{\mathbf{r}} \times \dot{\hat{\mathbf{r}}} \quad [46]$$

where we have set $\hat{\mathbf{r}} \cdot \hat{\mathbf{r}}_T = 1$. So with the generic relation between derivatives of a vector relative to body and space coordinates described above with $\hat{\mathbf{r}} = \cos f \mathbf{e}_1 + \sin f \mathbf{e}_2$ and $\dot{\hat{\psi}} = \dot{\psi} (\sin i \sin \Omega' \mathbf{e}_1 - \sin i \cos \Omega' \mathbf{e}_2 + \cos i \mathbf{e}_3) + \dot{\Omega}' \mathbf{e}_3$, it is easy to obtain

$$\begin{aligned} \hat{\mathbf{r}} \times \dot{\hat{\mathbf{r}}} &= \dot{\psi} \sin i \cos(f - \Omega') (-\sin f \mathbf{e}_1 + \cos f \mathbf{e}_2) \\ &\quad + (\dot{f} - \dot{\Omega}' - \dot{\psi} \cos i) \mathbf{e}_3 \end{aligned} \quad [47]$$

where i and Ω' are inclination and longitude of the ascending node of the $x'y'$ (equator) plane on the xy (orbit) plane. Eqn [47] is substituted into eqn [46] and averaged over the orbit period. Useful averages are

$$\begin{aligned} \left\langle \frac{a^6}{r^6} \dot{f} \right\rangle &= n \left(1 + \frac{15}{2} e^2 + \frac{45}{8} e^4 + \frac{5}{16} e^6 \right) / (1 - e^2)^6 \\ &= nf_1(e) \\ \left\langle \frac{a^6}{r^6} \right\rangle &= \left(1 + 3e^2 + \frac{3}{8} e^4 \right) / (1 - e^2)^{9/2} = f_2(e) \\ \left\langle \frac{a^6}{r^6} \cos 2f \right\rangle &= \left(\frac{3}{2} e^2 + \frac{1}{4} e^4 \right) / (1 - e^2)^{9/2} = f_3(e) \\ \left\langle \frac{a^6}{r^6} \cos^2 f \right\rangle &= \frac{f_2(e) + f_3(e)}{2} \\ \left\langle \frac{a^6}{r^6} \sin^2 f \right\rangle &= \frac{f_2(e) - f_3(e)}{2} \end{aligned} \quad [48]$$

where $\dot{f} = n(a^2 \sqrt{1 - e^2}) / r^2$ has been used in the first of eqns [48]. The averaged tidal torque on m is thus

$$\begin{aligned} \langle \mathbf{T}_T \rangle &= \frac{3k_2 G m_0^2 R^5 \Delta t}{a^6} \left\{ -\dot{\psi} \sin i \sin \Omega' \left(\frac{f_2(e) - f_3(e)}{2} \right) \mathbf{e}_1 \right. \\ &\quad + \dot{\psi} \sin i \cos \Omega' \left(\frac{f_2(e) + f_3(e)}{2} \right) \mathbf{e}_2 \\ &\quad \left. + [nf_1(e) - f_2(e)\dot{\psi} \cos i] \mathbf{e}_3 \right\} \end{aligned} \quad [49]$$

where we have neglected $\dot{\Omega}'$ compared to n and $\dot{\psi}$.

As an example, consider the retardation of the rotation of a satellite in a circular ($e=0$), equatorial ($i=0$) orbit from the tidal torque. The time rate of change of the angular momentum of the satellite $C(d\psi/dt) = \langle \mathbf{T}_T \rangle$, where C is the moment of inertia of m about its spin axis and where only the \mathbf{e}_3 term remains in eqn [49]. So

$$\begin{aligned} \frac{d\psi}{dt} &= -\frac{3k_2 G m_0^2 R^5}{Ca^6} \Delta t (\dot{\psi} - n) \\ &= -\frac{15}{2} k_2 \frac{m_0}{m} \frac{R^3}{a^3} \frac{n}{Q_0} (\dot{\psi} - n) \end{aligned} \quad [50]$$

where $\Delta t = 1/(Q_0 n)$, $C = 2mR^2/5$, and $n^2 = Gm_0/a^3$ have been used. The expression for Δt follows from the discussion below with Q_0 being the dissipation function at frequency n . The solution of eqn [50] is a

decaying exponential for $\dot{\psi} - n$ (if n is nearly constant) with time constant

$$\tau_T = \frac{2}{15} \frac{1}{k_2} \frac{m}{m_0} \frac{a^3}{R^3} \frac{Q_0}{n} \quad [51]$$

With $k_2 = 0.05$ and $Q_0 = 100$, $\tau_T \approx 4300$ years for Ganymede and 57 000 years for Callisto at their current distances from Jupiter. Note that for $e=0$ the asymptotic tidal state is $\dot{\psi} = n$, (i.e., rotation synchronous with the orbital mean motion), but for $e \neq 0$, the asymptotic tidal state is $(\dot{\psi} > n)$ determined by the vanishing of $\langle T_T \rangle$ in eqn [49]. Satellites with $e \neq 0$ are able to maintain synchronous rotation because the torque on the permanent axial asymmetry dominates $\langle T_T \rangle$ and forces the axis of minimum moment of inertia to librate about the direction to the primary whenever the satellite is at perapse.

The relationship between the dissipation function Q and Δt follows from a simple example. The dissipation parameter Q for a system oscillating at frequency ω is defined by (e.g., Lambeck, 1980 p. 14)

$$\frac{1}{Q} = \frac{\oint \frac{dE}{dt} dt}{2\pi E^*} = \omega \Delta t \quad [52]$$

where the numerator is the energy dissipated during a complete cycle of oscillation and E^* is the maximum energy stored during the cycle. For a tidally distorted, nearly spherical body with the disturbing body in a circular equatorial orbit, a cycle would consist of half a rotation of the distorted body relative to the body causing the tide. For a complex tide generated by a noncircular, nonequatorial orbit of the disturbing body, each periodic term in a Fourier decomposition of the tide would have its own maximum stored energy and dissipation over a complete cycle of oscillation.

The response of an oscillator with forcing function $F = A' \sin \omega t$ when $\omega \ll \omega_0$, with ω_0 being the lowest frequency of free oscillation, can be represented by $x = B' \sin \omega(t - \Delta t)$, where Δt is the phase lag in the response due to the dissipation as was assumed above for the tidal response. The rate at which the forcing function does work is $dE/dt = F\dot{x} = A' B' \omega \sin \omega t \cos \omega(t - \Delta t)$. Then

$$\begin{aligned} E(t) &= \int_{t_1}^t F\dot{x} dt \\ &= A' B' \omega \left[\frac{-\cos(2\omega t - \omega\Delta t)}{4\omega} + \frac{\sin \omega \Delta t}{2} t \right]_{t_1}^t \end{aligned} \quad [53]$$

The first term in eqn [53] is the periodic storage of energy in the oscillator, and the second is the secular loss of energy. The maximum energy stored is just twice

the coefficient of the cosine term, $E^* = A' B' / 2$, and the energy dissipated during a complete period of $2\pi/\omega$ is $\Delta E = A' B' \pi \omega \Delta t$ with $\sin \omega \Delta t \approx \omega \Delta t$. We use the energy stored as the energy increment above the minimum energy in the first term since the stored tidal energy will always be an increase over the relaxed spherical shape of the body. Hence $1/Q = \omega \Delta t$ as indicated in eqn [52]. Since Δt is independent of frequency, Q is inversely proportional to frequency in this model.

The tidal effective Q of the solid Earth has been estimated from satellite observations to be 280 ± 70 (Ray *et al.*, 2001) which is close to the value obtained by the attenuation of normal modes of the Earth (Widmer *et al.*, 1991). For Mars, $Q \approx 100$ from the secular acceleration of the satellite Phobos. (Smith and Born, 1976; Bills *et al.*, 2005). For Mars the dominant frequency is $\omega = 2(\dot{\psi} - n_p)$, where $\dot{\psi}$ is Mars rotational angular velocity and n_p is Phobos' mean orbital motion. The factor 2 follows from there being two tidal cycles for each synodic period. With a rotation period of $24^h 37^m 23^s$ and an orbital period of 0.319 days, $\omega = 3.14 \times 10^{-4}$ rad s $^{-1}$ leading to $\Delta t \approx 32$ s for $Q = 100$. For Mercury, the fundamental tidal period is the orbit period, since Mercury rotates 180° relative to the Sun for each orbit. For the attenuation of seismic waves in the Earth and from laboratory experiments on Earth-like materials, Q is proportional to frequency raised to a small, positive fractional (0.2–0.4) power with nominal values near $Q = 100$ (e.g., Shito *et al.*, 2004), but to avoid an infinite discontinuity in the torque whenever a tidal frequency passes through zero while changing sign, it is not unlikely that something close to Q being inversely proportional to frequency will prevail for the small tidal frequencies. Although one might infer that Q is inversely proportional to frequency near a particular frequency to smooth the transition from positive to negative frequency, it cannot be the case that the same proportionality constant will apply over a very wide range of frequency as indicated by $Q \propto \omega^{0.2-0.4}$ as obtained by Shito *et al.* (2004). We can only write $\Delta t = P' / (2\pi Q_0)$, with Q_0 being the value appropriate to a small range of frequencies near the dominant tidal frequency of the problem at hand (e.g., near the frequency of $2\pi/(88$ days) for Mercury).

For energy dissipation within a satellite, the approach of Peale and Cassen (1978) is simplified by Wisdom (2004). The rate at which work is done on a satellite by the tide is given by

$$\frac{dE}{dt} = - \int_{\text{body}} \rho \mathbf{v} \cdot \nabla W_2 dV \quad [54]$$

where W_2 is the above tide raising potential, \mathbf{v} is the velocity of a volume element dV of density ρ . If we can assume the satellite is incompressible, then $\nabla \cdot \mathbf{v} = 0$, and from $\nabla \cdot W_2 \mathbf{v} = \mathbf{v} \cdot \nabla W_2 + W_2 \nabla \cdot \mathbf{v}$, we can write

$$\begin{aligned}\frac{dE}{dt} &= - \int_{\text{body}} \rho \nabla \cdot (W_2 \mathbf{v}) dV \\ &= -\rho \int_{\text{surface}} W_2 \mathbf{v} \cdot \mathbf{n} dS\end{aligned}\quad [55]$$

where the last form follows from Gauss's law. In eqn [55] \mathbf{n} is the normal to a surface element dS , and we have taken ρ outside the integral by assuming a uniform density.

The rate at which the height of the surface changes is $\mathbf{v} \cdot \mathbf{n} = d\Delta R/dt \approx b_2 dW'_2/dt/g$ (eqn [35]), where $b_2 = 5k_2/3$ is the second degree displacement Love number (Love, 1944), and g is the surface gravity acceleration of the satellite. The prime on W_2 indicates the time delay in the response of the body to the forcing function. Then

$$\frac{dE}{dt} = -\frac{\rho b_2}{g} \int_{\text{surface}} W_2 \frac{d}{dt}(W'_2) dS\quad [56]$$

If higher-order terms are necessary in the tidal potential for close satellites, eqn [56] would be a sum of terms involving the Love numbers b_3 , b_4 , etc.

If we consider the case where the satellite equator and orbit planes are coincident, $\mathbf{r} = (r \cos f, r \sin f, 0)$, and $\mathbf{R} = (R \cos(\phi + \dot{\psi}t), R \sin(\phi + \dot{\psi}t), 0)$, where R , ϕ are the ordinary spherical radial and azimuthal coordinates of a point on the surface in a coordinate system fixed in the satellite, and the $\dot{\psi}t$ accounts for the rotation of the satellite relative to an inertial system. For small eccentricities, we can write $a/r \approx 1 + e \cos nt$, $\cos f \approx \cos nt + e(\cos 2nt - 1)$ and $\sin f = \sin nt + e \sin 2nt$ to first order in eccentricity e . For the time-lagged potential W'_2 , t is replaced by $t - \Delta t$, and for a satellite rotating synchronously with its orbital motion, $\dot{\psi} = n$. With these substitutions, the average rate of energy dissipation is found by performing the surface integral and averaging over an orbit period.

$$\begin{aligned}\frac{dE}{dt} &= -\frac{42}{5} \frac{\pi \rho b_2}{g} \frac{G^2 m_p^2 R^6}{a^6} e^2 n \sin n\Delta t \\ &\approx -\frac{42}{19} \frac{\pi \rho^2 n^5 R^7 e^2}{\mu Q}\end{aligned}\quad [57]$$

where $b_2 \approx 5\rho g R / (19\mu)$ from the definition of k_2 , $n^2 = Gm_p/a^3$ has been used, and $\sin n\Delta t \rightarrow n\Delta t = 1/Q$. The last form in eqn [57] agrees with that in Peale

et al. (1979) as corrected in Peale (2003). Equation [57] is the lowest-order approximation for the tidal dissipation in a synchronously rotating satellite with orbital eccentricity e . For very large eccentricities, higher-order terms must be included.

The tidal dissipation in a synchronously rotating satellite will tend to reduce the eccentricity as $e \neq 0$ is the cause of the dissipation. The spin angular momentum of the satellite is conserved because of the synchronous lock. The specific orbital angular momentum $[G(m_p + m_s)a(1 - e^2)]^{1/2}$ cannot thus gain angular momentum from the satellite spin. The orbital energy $-Gm_p m_s / 2a$ must decrease if energy is dissipated in the satellite, and a must thereby decrease. But the conserved angular momentum means e must decrease if a decreases. At the same time, the tide raised on the planet by the satellite tends to increase the eccentricity. The greater tidal force at the periape tends to fling the satellite to a greater apoapse distance than it would have reached without the kick, thereby increasing e . The variation in the eccentricity from the two effects is (Goldreich, 1963)

$$\begin{aligned}\frac{de}{dt} &= \frac{57}{8} k_{2p} n \frac{m_s}{m_p} \frac{R_p^5}{a^5} \frac{e}{Q_p} \text{sign}(\dot{\psi}_p - n_s) \\ &\quad - \frac{21}{2} k_{2s} n \frac{m_p}{m_s} \frac{R_s^5}{a^5} \frac{e}{Q_s}\end{aligned}\quad [58]$$

where $\dot{\psi}_p$ and n_s are the spin angular velocity and orbital mean motion, respectively, and subscripts p and s refer to planet and satellite. Whether eccentricity of a satellite increases or decreases as the orbit evolves under tidal evolution depends on the relative magnitude of the two terms in eqn [58].

The reader should be aware of the uncertainty and somewhat arbitrariness of tidal models. Our analysis of tidal dissipation using a constant time lag in the response of a planet or satellite to the forcing function leads to $Q \propto 1/\omega$, where ω is a tidal frequency, is the simplest. It has the most straightforward development, while yielding relatively intuitive tidal evolution. Other models frequently used include the Maxwell solid model (e.g., Hussmann and Spohn, 2004) where $Q \propto \omega$ for high frequencies and $Q \propto 1/\omega$ at very low frequencies, and the $Q = \text{constant}$ model (e.g., Kaula, 1964). A modification of the Maxwell solid model might possibly be constructed to yield the empirical $Q \propto \omega^\alpha$ at high frequencies and $1/\omega$ at low frequencies, where α is the small fractional exponent discussed by Shito *et al.* (2004), but such an attempt has not been made.

We have seen above that nominal values of Q (or Q_0 in the model detailed above) are near 100 for Mars

and for the damping of seismic waves in the Earth's mantle, and near 300 for satellite determinations of the solid Earth response to the Moon and for the damping of seismically excited normal modes of the Earth. $Q=100$ is an often adopted value representing Earth-like materials. However, the damping of seismic waves in the Moon's upper layers leads to $Q \approx 4000$ at 4 Hz and $Q \approx 7000$ at 8 Hz (Nakamura and Koyama, 1982), but the monthly tidal $Q \approx 37$ and the annual $Q \approx 60$ (Williams *et al.*, 2001). The widely varying values of Q reflect the nature of the materials where the dissipation is taking place, where the dryness of the surface layers of the Moon may lead to the high values of Q , and partial melts in the interior may yield the low tidal Q s. Finally, a totally different procedure is needed for the unsolved problem of determining effective Q s of gaseous planets (and stars), where resonances of tidal frequencies with normal modes of oscillation of the gaseous body become important. In summary, the uncertainty of the proper tidal model to use in various circumstances persists, but the simplest tidal theory described above can be used to determine qualitative constraints on the origin of the natural satellites from their inferred evolutions.

References

- Agnor CB, Canup RM, and Levison HF (1999) On the character and consequences of large impacts in the late stage of terrestrial planet formation. *Icarus* 100: 307–325.
- Agnor CB and Hamilton DP (2004) Satellite capture via binary exchange reactions: Application to Triton. *Bulletin of American Astronomical Society* 36: 1169.
- Agnor CB and Hamilton DP (2006) Neptune's capture of its Moon Triton in a binary-planet gravitational encounter. *Nature* 441: 192–194.
- Alfvén H (1963) The early history of the Moon and the Earth. *Icarus* 1: 357–363.
- Alfvén H and Arrhenius G (1969) Two alternatives for the history of the Moon. *Science* 165: 11–17.
- Alfvén H and Arrhenius G (1972) Origin and evolution of the Earth–Moon system. *The Moon* 5: 210–230.
- Alfvén H and Arrhenius G (1976) *Evolution of the Solar System*. NASA, Washington, DC: Scientific and Technical Information Office: NASA SP-345.
- Alibert Y, Mordasini C, Benz W, and Winisdoerffer C (2005a) Models of giant planet formation with migration and disc evolution. *Astronomy and Astrophysics* 434: 343–353.
- Alibert Y and Mousis O (2006) Structure and evolution of Saturn's subnebula – Implications for the formation of Titan. *Lunar Planetary Science XXXVII*, Abstract 1141.
- Alibert Y, Mousis O, and Benz W (2005b) Modeling the Jovian subnebula. I: Thermodynamic conditions and migration of proto-satellites. *Astronomy and Astrophysics* 439: 1205–1213.
- Anderson JD, Schubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998a) Distribution of rock, metals and ice in Callisto. *Science* 280: 1573–1576.
- Anderson JD, Schubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998b) Europa's differentiated internal structure: Inferences from four Galileo encounters. *Science* 281: 2019–2022.
- Anderson JD, Jacobson RA, McElrath TP, Moore WB, Schubert G, and Thomas PC (2001) Shape, mean radius, gravity field and interior structure of Callisto. *Icarus* 153: 157–161.
- Balbus S (2003) Enhanced angular momentum transport in accretion disks. *Annual Review of Astronomy and Astrophysics* 41: 555–597.
- Banfill D and Murray N (1992) A dynamical history of the inner Neptunian satellites. *Icarus* 99: 390–401.
- Bate MR, Lubow SH, Ogilvie GI, and Miller KA (2003) Three-dimensional calculations of high- and low-mass planets embedded in protoplanetary discs. *Monthly Notices of the Royal Astronomical Society* 341: 213–229.
- Bell KR and Lin DNC (1994) Using Fu Orionis outbursts to constrain self-regulated protostellar disk models. *The Astrophysical Journal* 474: 987–1004.
- Benner LAM and McKinnon WB (1995) Orbital behavior of captured satellites: The effect of solar gravity on Triton's postcapture orbit. *Icarus* 114: 1–20.
- Benz W, Slattery WL, and Cameron AGW (1986) The origin of the Moon and the single impact hypothesis. I. *Icarus* 66: 515–535.
- Benz W, Slattery WL, and Cameron AGW (1987) The origin of the Moon and the single-impact hypothesis. II. *Icarus* 71: 30–45.
- Bills BG, Neumann GA, Smith DE, and Zuber MT (2005) Improved estimate of tidal dissipation within Mars from MOLA observations of the shadow of Phobos. *Journal of Geophysical Research* 110: E07004.
- Borderies N, Goldreich P, and Tremaine S (1984) Excitation of inclinations in ring-satellite systems. *The Astrophysical Journal* 284: 429–434.
- Boss AP and Peale SJ (1986) Dynamical constraints on the origin of the Moon. In: Hartman WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 59–101. Huston, TX: Lunar and Planetary Institute.
- Brunini A (2006a) Origin of the obliquities of the giant planets in mutual interactions in the early solar system. *Nature* 440: 1163–1165.
- Brunini A (2006b) Retraction: Origin of the obliquities of the giant planets in mutual interactions in the early solar system. *Nature* 443: 1013 (letters to editor).
- Buie MW, Grundy WM, Young EF, Young LA, and Stern SA (2006) Orbits and photometry of Pluto's satellites: Charon, S/2005 P1 and S/2005 P2. *The Astrophysical Journal* 132: 290–298.
- Burns JA (1986) Some background on satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 1–38. Tucson AZ: University of Arizona Press.
- Burns JA (2003) Smashed into orbit, a review of *The Big Splat, or How Our Moon Came to Be*, by Dana MacKenzie. *Nature* 425: 664–665.
- Bryden G, Chen X, Lin DNC, Nelson RP, and Papaloizou JCB (1999) Tidally induced gap formation in protostellar disks: Gap clearing and suppression of protoplanetary growth. *The Astrophysical Journal* 514: 344–367.
- Cameron AGW and Ward WR (1976) The origin of the Moon (abstract). In: *Lunar Science VII*, pp. 120–122. Houston, AZ: The Lunar Science Institute.
- Canup RM (2004a) Dynamics of lunar formation. *Annual Review of Astronomy and Astrophysics* 42: 441–475.
- Canup RM (2004b) Simulations of a late Lunar-forming impact. *Icarus* 168: 433–456.
- Canup RM (2005) A giant impact origin of Pluto–Charon. *Science* 307: 546–550.

- Canup RM and Asphaug E (2001) Origin of the Moon in a giant impact near the end of the Earth's formation. *Nature* 412: 708–712.
- Canup RM and Esposito LW (1996) Accretion of the Moon from an impact-generated disk. *Icarus* 119: 427–446.
- Canup RM and Righter K (2000) *Origin of the Earth and Moon*. Tucson, AZ: University of Arizona Press.
- Canup RM and Ward WR (2000) A possible impact origin of the Uranian satellite system. *Bulletin of the American Astronomical Society* 32: 1105.
- Canup R and Ward WR (2001) Steady-state conditions in a Jovian protosatellite accretion disk. Paper presented at the Jupiter Conference, Boulder, CO, 25–30 Jun 2001.
- Canup R and Ward WR (2002) Formation of the Galilean satellites: Conditions of accretion. *The Astrophysical Journal* 124: 3404–3423.
- Canup R and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441: 834–839.
- Carlson RW (1999) A tenuous carbon dioxide atmosphere on Jupiter's moon Callisto. *Science* 283: 820–821.
- Carruba V, Burns JA, Nicholson PD, and Gladman BJ (2001) On the inclination distribution of the Jovian irregular satellites. *Icarus* 158: 434–449.
- Carruba V, Burns JA, Nicholson PD, and Gladman BJ (2002) On the inclination distribution of the Jovian irregular satellites. *Icarus* 158: 434–439.
- Carruba V, Nesvorný D, Burns JA, and Čuk M (2003) Kozai resonators among distant moons of the Jovian planets. *Bulletin of American Astronomical Society* 35: 1044.
- Carruba V, Nesvorný D, Burns JA, Čuk M, and Tsiganis K (2004) Chaos and the effects of planetary migration on the orbit of S/2000 S5 Kiviuq. *The Astrophysical Journal* 128: 1899–1915.
- Chambers JE (2001) Making more terrestrial planets. *Icarus* 152: 205–224.
- Christensen EJ, Born GH, Hildebrand CE and Williams BG (1977) The name of phobos from Viking flybys. *Geophysical Research Letters* 4: 555–557.
- Colombo G and Franklin FA (1971) On the formation of the outer satellite groups of Jupiter. *Icarus* 15: 186–189.
- Colwell JE and Esposito LW (1992) Origins of the rings of Uranus and Neptune. I: Statistics of satellite disruptions. *Journal of geophysical Research* 97E: 10227–10241.
- Conway BA (1982) On the history of the lunar orbit. *Icarus* 51: 610–622.
- Craddock RA (1994) The origin of Phobos and Deimos. *Lunar and Planetary Science Conference XXV*: 293–294.
- Čuk M and Burns JA (2004) Gas-drag-assisted capture of Himalia's family. *Icarus* 167: 369–381.
- Čuk M and Gladman BJ (2005) Constraints on the orbital evolution of Triton. *The Astrophysical Journal* 626: L113–L116.
- D'Angelo G, Henning T, and Kley W (2002) Nested-grid calculations of disk–planet interaction. *Astronomy and Astrophysics* 385: 647–670.
- Darwin GH (1879) On the precession of a viscous spheroid, and on the remote history of the Earth. *Philosophical Transactions of the Royal Society* 170: 447–530.
- Darwin GH (1980) On the secular changes in the elements of the orbit of a satellite revolving about a tidally distorted planet. *Philosophical Transactions of the Royal Society* 171: 713–891.
- Dickey JO, Bender PL, Faller JE, et al. (1994) Lunar laser ranging: A continuing legacy of the Apollo program. *Science* 265: 482–490.
- Dobrovolskis AR (1982) Internal stresses in Phobos and other triaxial bodies. *Icarus* 52: 136–148.
- Dobrovolskis AR, Peale SJ, and Harris AW (1997) The dynamics of the Pluto–Charon binary. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 159–189. Tucson AZ: University of Arizona Press.
- Dones L and Tremaine S (1993) On the origin of planetary spins. *Icarus* 103: 67–92.
- Duncan MJ, Levison HF, and Lee MH (1998) A multiple time step symplectic algorithm for integrating close encounters. *Astronomical Journal* 116: 2067–2077.
- Duxbury TC and Callahan JD (1981) Pole and prime meridian expressions for PHOBOS and Deimos. *Astronomical Journal* 86: 1722–1727.
- Duxbury TC and Veverka J (1978) Deimos encounter by Viking – Preliminary imaging results. *Science* 201: 812–814.
- Dycemons V (1978) Formation of planetary and satellite systems. *Astrophysics and Space Sciences* 58: 521–523.
- Estrada PR and Mosqueira I (2006) A gas-poor planetesimal capture model for the formation of giant planet satellite systems. *Icarus* 181: 486–509.
- Farinella P, Milani A, Nobili AM, and Valsecchi GB (1980) Some remarks on the capture of Triton and the origin of Pluto. *Icarus* 44: 810–812.
- Folkner WM, Yoder CF, Yuan DN, Standish EM, and Preston RA (1997) Interior structure and seasonal mass redistribution of Mars from radio tracking of Mars pathfinder. *Science* 278: 1749–1752.
- Garrick-Bethell I, Wisdom J, and Zuber MT (2006) Evidence for a past high eccentricity lunar orbit. *Science* 313: 652–655.
- Gerstenkorn H (1955) Über gezeitenreibung beim zweikörperproblem. *Zeitschrift für Astrophysik* 36: 245–274.
- Gerstenkorn H (1969) The earliest past of the Earth–Moon system. *Icarus* 11: 189–207.
- Gladman BJ, Kavelaars JJ, Holman M, et al. (2000) The discovery of Uranus XIX, XX, and XXI. *Icarus* 147: 320–324 (erratum (2000) 148: 320).
- Gladman B, Kavelaars JJ, Holman M, et al. (2001) Discovery of 12 satellites of Saturn exhibiting orbital clustering. *Nature* 412: 163–169.
- Gladman BJ, Nicholson PD, Burns JA, et al. (1998) Discovery of two distant irregular moons of Uranus. *Nature* 392: 897–899.
- Goldreich P (1963) On the eccentricity of satellite orbits in the solar system. *Monthly Notices of the Royal Astronomical Society* 126: 257–268.
- Goldreich P (1965) Inclination of satellites orbits about an oblate processing planet. *Astronomical Journal* 70: 5–9.
- Goldreich P (1966) History of the lunar orbit. *Reviews of Geophysics* 4: 411–439.
- Goldreich P, Lithwick Y, and Sari R (2002) Formation of Kuiper-belt binaries by dynamical friction and three-body encounters. *Nature* 420: 643–646.
- Goldreich P, Murray N, Longaretti PY, and Banfield D (1989) Neptune's story. *Science* 245: 500–504.
- Goldreich P and Soter S (1966) Q in the solar system. *Icarus* 5: 375–389.
- Goldreich P and Ward WR (1973) The formation of planetesimals. *The Astrophysical Journal* 183: 1051–1062.
- Greenzweig Y and Lissauer JJ (1990) Accretion rates of protoplanets. *Icarus* 87: 40–77.
- Greenzweig Y and Lissauer JJ (1991) Accretion rates of protoplanets II: Gaussian distributions of planetesimal velocities. *Icarus* 100: 440–463.
- Greenzweig Y and Lissauer JJ (1992) Accretion rates of protoplanets. II: Gaussian distributions of planetesimal velocities. *Icarus* 100: 440–463.
- Halliday AH (2004) Mixing, volatile loss and compositional change during impact driven accretion of the Earth. *Nature* 427: 505–509.
- Hamilton DP and Burns JA (1991) Orbital stability zones about asteroids. *Icarus* 92: 118–131.

- Hamilton DP and Krivov AV (1997) Dynamics of distant moons of asteroids. *Icarus* 128: 241–249.
- Harris AW (1977) The effect of tidal friction on the origin and thermal evolution of the Moon (abstract). In: *Lunar Science VIII*, pp 401–402. Houston AZ: The Lunar Science Institute.
- Harris AW (1978) Satellite formation, II. *Icarus* 34: 128–145.
- Harris AW and Kaula WM (1975) A co-accretion model of satellite formation. *Icarus* 24: 516–524.
- Hartmann WK and Davis DR (1975) Satellite-sized planetesimals and lunar origin. *Icarus* 24: 504–515.
- Hartmann WK, Phillips RJ, and Taylor GJ (1986) *Origin of the Moon*. Houston, TX: Lunar and Planetary Institute.
- Heppenheimer TA and Porco CC (1977) New contributions to the problem of capture. *Icarus* 30: 385–401.
- Hibbitts CA, Klemaszewski JE, McCord TB, Hansen GB, and Greeley R (2002) CO₂-rich impact craters on Callisto. *Journal of Geophysical Research (Planets)* 107(14): 1–12.
- Hibbitts CA, McCord TB, and Hansen GB (2000) Distributions of CO₂ and SO₂ on the surface of Callisto. *Journal of Geophysical Research* 430: 22541–22558.
- Hibbitts CA, Pappalardo RT, Hansen GB, and McCord TB (2003) Carbon dioxide on Ganymede. *Journal of Geophysical Research (Planets)* 108(2): 1–22.
- Hollenbach DJ, York HW, and Johnstone D (2000) Disk dispersal around young stars. In: Manning V, Boss AP, and Russell SS (eds.) *Protostars and Planets*, pp. 401. Tucson, AZ: University of Arizona Press.
- Holman MJ, Kavelaars JJ, Grav T, Gladman BJ, et al. (2004) Discovery of five irregular moons of Neptune. *Nature* 430: 865–867.
- Horedt GP (1976) Pre-capture orbits of the Moon. *The Moon* 15: 439–443.
- Hussmann H and Spohn T (2004) Thermal-orbital evolution of Io and Europa. *Icarus* 171: 391–410.
- Ida S, Canup RM, and Stewart GR (1997) Lunar accretion from an impact generated disk. *Nature* 389: 353–357.
- Imaeda Y and Inutsuka S (2002) Shear flows in smoothed particle hydrodynamics. *Astronomy and Physics Journal* 569: 501–518.
- Jacobsen SB (2005) The Hf–W isotopic system and the origin of the Earth and Moon. *Annual Review of Earth and Planetary Sciences* 33: 531–570.
- Jones JH and Palme H (2000) Geochemical constraints on the origin of the Moon. In: Canup RM and Richter J (eds.) *Origin of the Earth and Moon*, pp. 197–216. Tucson, AZ: University of Arizona Press.
- Kary DM and Lissauer JJ (1995) Nebular gas drag and planetary accretion. II. Planet on an eccentric orbit. *Icarus* 117: 1–24.
- Kary DM, Lissauer JJ, and Greenzweig Y (1993) Nebular gas drag and planetary accretion. *Icarus* 106: 288–307.
- Kaula WM (1964) Tidal dissipation by solid friction and the resulting orbital evolution. *Reviews of Geophysics* 2: 661–685.
- Kivelson MG, Khurana KK, Coroniti FV, et al. (1997) Magnetic field and magnetosphere of Ganymede. *Geophysical Research Letters* 24: 2155–2158.
- Kivelson MG, Khurana KK, Stevenson DJ, et al. (1999) Europa and Callisto: Induced or intrinsic fields in a periodically varying plasma environment. *Journal of Geophysical Research* 104: 4609–4626.
- Kivelson MG, Khurana KK, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.
- Kleine T, Mezger K, Palme H, and Münker C (2004) The W isotope evolution of the bulk silicate Earth: Constraints on the timing and mechanisms of core formation and accretion. *Earth and Planetary Science Letters* 228: 109–123.
- Kleine T, Palme H, Mezger K, and Halliday AN (2005) Hf–W chronometry of lunar metals and the age and early differentiation of the Moon. *Science* 310: 1671–1674.
- Kley W (1999) Mass flow and accretion through gaps in accretion discs. *Monthly Notices of the Royal Astronomical Society* 303: 696–710.
- Kley W, D’Angelo G, and Henning T (2001) Three-dimensional simulations of a planet embedded in a protoplanetary disk. *The Astrophysical Journal* 547: 457–464.
- Kokubo E, Ida S, and Makino J (2000) Evolution of a circumterrestrial disk and formation of a single Moon. *Icarus* 148: 419–436.
- Kolyuka YF, Efimov AE, Kudryavtsev SM, Margorin OR, Tarasov VP, and Tikhonov VF (1990) Refinement of gravitational constant of PHOBOS from PHOBOS-2 tracking data. *Soviet Astronomy Letters* 16: 168–170.
- Kuskov OL and Kronrod VA (2005) Internal structure of Europa and Callisto. *Icarus* 177: 550–569.
- Lambeck K (1980) *The Earth’s Variable Rotation: Geophysical Causes and Consequences*. New York: Cambridge University Press.
- Laskar J and Robutel P (1993) The chaotic obliquity of planets. *Nature* 361: 608–612.
- Lee MH and Peale SJ (2000) Making Hyperion. *Bulletin of American Astronomical Society* 32: 1078.
- Lee MH and Peale SJ (2006) On the orbits and masses of the satellites of the Pluto–Charon system. *Icarus* (submitted).
- Lee MH, Peale SJ, Pfahl E, and Ward WR (2006) Evolution of the obliquities of the giant planets during migration, DPS abstract. *Bulletin of American Astronomical Society* 38: 611.
- Lee MH, Peale SJ, Pfahl E and Ward WR (2007) Evolution of the obliquities of the giant planets during migration. *Icarus* (in press.)
- Lin DNC and Papaloizou J (1985) On the dynamical origin of the solar system. In: Black D and Matthews MS (eds.) *Protostars and Planets II*, pp. 981–1072. Tucson, AZ: University of Arizona Press.
- Lissauer JJ (1987) Timescales for planetary accretion and structure of the protoplanetary disk. *Icarus* 69: 249–265.
- Lissauer JJ (1995) Urey Prize lecture: On the diversity of plausible planetary systems. *Icarus* 114: 217–236.
- Love AEH (1944) *A Treatise on the Mathematical Theory of Elasticity*. New York: Dover Publication.
- Lubow SH, Seibert M, and Artymowicz P (1999) Disk accretion onto high-mass planets. *The Astrophysical Journal* 526: 1001–1012.
- Lunine JI and Nolan MC (1992) A massive early atmosphere on Triton. *Icarus* 100: 221–234.
- Lunine JI and Stevenson DJ (1982) Formation of the galilean satellites in a gaseous nebula. *Icarus* 52: 14–39.
- Lynden-Bell D and Pringle JE (1974) The evolution of viscous disks and the origin of the nebular variables. *Monthly Notices of the Royal Astronomical Society* 168: 603–637.
- Lyttleton RA (1967) Dynamical capture of the Moon by the Earth. *Proceedings of the Royal Society of London A296:* 285–292.
- MacKenzie D (2003) *The Big Splat, or How Our Moon Came to Be*. Hoboken, NJ: John Wiley.
- Margot J-L, Nolan MC, Benner LAM, et al. (2002) Binary asteroids in the near-earth object population. *Science* 296: 1445–1448.
- McCord TB, Hansen GB, Clark RN, et al. (1998) Non-water-ice constituents in the surface material of the icy Galilean satellites from the Galileo near-infrared mapping spectrometer investigation. *Journal of Geophysical Research* 103: 8603–8626.
- McKinnon WB (1984) On the origin of Triton and Pluto. *Nature* 311: 355–358.
- McKinnon WB (1989) On the origin of the Pluto–Charon binary. *The Astrophysical Journal* 344: L41–L44.
- McKinnon WB (1997) Mystery of Callisto: Is it undifferentiated? *Icarus* 130: 540–543.

- McKinnon WB (2006) On convection in ice I shells of outer solar system bodies, with detailed application to Callisto. *Icarus* 183: 435–450.
- McKinnon WB and Leith AC (1995) Gas drag and the orbital evolution of a captured Triton. *Icarus* 118: 392–413.
- McKinnon WB, Lunine JI, and Banfield D (1995) Origin and evolution of Triton. In: Cruikshank DP (ed.) *Neptune and Pluto*, pp. 807–877. Tucson, AZ: University of Arizona Press.
- Merline WJ, Weidenschilling SJ, Durda DD, Margot J-L, Pravec P, and Storrs AD (2002) Asteroids do have satellites. In: Bottke Jr. WF, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, pp. 289–312. Tucson: University of Arizona Press.
- Mignard F (1981a) The lunar orbit revisited, III. *Moon and Planets* 24: 189–207.
- Mignard F (1981b) Evolution of Martian satellites. *Monthly Nature of the Royal Astronomical Society* 194: 365–379.
- Mosqueira I and Estrada PR (2003a) Formation of the regular satellites of giant planets in an extended gaseous nebula. I: Subnebula model and accretion of satellites. *Icarus* 163: 198–231.
- Mosqueira I and Estrada PR (2003b) Formation of the regular satellites of giant planets in an extended gaseous nebula. II: Satellite migration and survival. *Icarus* 163: 232–255.
- Mousis O and Alibert Y (2006) Modeling the Jovian subnebula. II: Composition of regular satellite ices. *Astronomy and Astrophysics* 448: 771–778.
- Mousis O, Pargamin J, Grasset O, and Sotin C (2002b) Experiments in the NH₃-H₂O system in the [0, 1 GPa] pressure range – Implications for the deep liquid layer of large icy satellites. *Geophysical Research Letters* 29(45): 1–4.
- Murray CD and Dermott SF (1999) *Solar System Dynamics*. Cambridge: Cambridge University Press.
- Nagel K, Breuer D, and Spohn T (2004) A model for the interior structure, evolution, and differentiation of Callisto. *Icarus* 169: 402–412.
- Nakamura Y and Koyama J (1982) Seismic Q of the lunar upper mantle. *Journal of Geophysical Research* 87: 4855–4861.
- Niemann HB, Atreya SK, Bauer SJ, et al. (2005) The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* 438: 779–784.
- Öpik EJ (1969) The Moon's surface. *Annual Review of Astronomy and Astrophysics* 7: 473–526.
- Öpik EJ (1971) Comments on lunar origin. *Irish Astronomical Journal* 10: 190–238.
- Pang KD, Pollack JB, Veverka J, Lane AL, and Ajello JM (1978) Multicolor Observations of PHOBOS with the Viking lander cameras – Evidence for a carbonaceous chondritic composition. *Science* 199: 66–69.
- Papaloizou JCB and Larwood JD (2000) On the orbital evolution and growth of protoplanets embedded in a gaseous disc. *Monthly Notices of the Royal Astronomical Society* 315: 823–833.
- Papaloizou JCB and Terquem C (1999) Critical protoplanetary core masses in protoplanetary disks and the formation of short-period giant planets. *The Astrophysical Journal* 521: 823–838.
- Peale SJ (1977) Rotation histories of the natural satellites. In: Burns J (ed.) *Planetary Satellites*, pp. 87–111. Tucson: University of Arizona Press.
- Peale SJ (1986) Orbital resonances, unusual configurations and exotic rotation states among planetary satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 159–223. Tucson: University of Arizona Press.
- Peale SJ (1993) The effect of the nebula on the Trojan precursors. *Icarus* 106: 308–322.
- Peale SJ (1999) Origin and evolution of the natural satellites. *Annual Review of Astronomy and Astrophysics* 37: 533–602.
- Peale SJ (2003) Tidally induced volcanism. *Celestial Mechanics and Dynamic Astronomy* 87: 129–155.
- Peale SJ and Cassen PM (1978) Contribution of tidal heating to lunar thermal history. *Icarus* 36: 245–269.
- Peale SJ, Cassen PM, and Reynolds RT (1979) Melting of Io by tidal dissipation. *Science* 203: 892–894.
- Peale SJ and Lee MH (2002) A primordial origin of the Laplace relation among the Galilean satellites. *Science* 298: 593–597.
- Pollack JB and Burns JA (1977) An origin by capture for the Martian satellites? *Bulletin of the American Astronomical Society* 9: 518.
- Pollack JB, Burns JA, and Tauber ME (1979) Gas drag in primordial circumplanetary envelopes: A mechanism for satellite capture. *Icarus* 37: 587–611.
- Pollack JB, Hollenbach D, Beckwith S, Simonelli DP, Roush T, and Fong W (1994) Composition and radiative properties of grains in molecular clouds and accretion disks. *The Astrophysical Journal* 421: 615–639.
- Pollack JB, Veveka J, Pang KD, Colburn DS, Lane AL, and Ajello JM (1978) Multicolor observations of PHOBOS with the Viking lander cameras – Evidence for a carbonaceous chondritic composition. *Science* 199: 66–69.
- Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active South Pole of Enceladus. *Science* 311: 1393–1401.
- Prandtl L (1952) *Essentials of Fluid Dynamics*, pp. 174–196. New York: Hafner.
- Pritchard ME and Stevenson DJ (2000) Thermal aspects of a lunar origin by giant impact. In: Canup RM and Richter K (eds.) *Origin of the Earth and Moon*, pp. 179–196. Tucson: University of Arizon Press.
- Ray R, Eanes RJ, and Lemoine FG (2001) Constraints on energy dissipation in the Earth's body tide from satellite tracking and altimetry. *Geophysical Journal International* 144: 471–480.
- Ringwood AE (1970) Origin of the Moon: The precipitation hypothesis. *Earth and Planetary Science Letters* 8: 131–140.
- Ringwood AE (1972) Some comparative aspects of lunar origin. *Physics of Earth and Planetary Interiors* 6: 366–376.
- Ruden SP and Pollack JB (1991) The dynamical evolution of the protosolar nebula. *The Astrophysical Journal* 375: 740–760.
- Ruskol EL (1961) The origin of the Moon I. *Soviet Astronomy-AJ* 4: 657–668.
- Ruskol EL (1963) The origin of the Moon II. *Soviet Astronomy-AJ* 7: 221–227.
- Ruskol EL (1972a) The origin of the Moon. III: Some aspects of the dynamics of the circumterrestrial swarm. *Soviet Astronomy-AJ* 15: 646–654.
- Ruskol EL (1972b) Possible differences in the chemical composition of the Earth and Moon, for a Moon formed in the circumterrestrial swarm. *Soviet Astronomy-AJ* 15: 1061–1063.
- Ruskol EL (1972c) On the initial distance of the Moon forming in the circumterrestrial swarm. *The Moon* 5: 402–404.
- Ruskol EL (1975) The origin of the Moon. In: *Proceedings of the Soviet-American Conference on Cosmochemistry of the Moon and Planets*, pp. 638–644, Nauka, Moscow (transl. in English in NASA SP-370 (1977), pp. 815–822).
- Safronov VS (1969) *Evolution of the Protoplanetary Cloud and the Formation of the Earth and Planets*. Moscow: Nauka Press, (NASA TT-677, 1972).
- Safronov VS, Pechernikova GV, Ruskol EL, and Vitiaz AV (1986) Protosatellite swarms. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 89–116. Tucson, AZ: University of Arizone Press.
- Saha P and Tremaine S (1993) The orbits of the retrograde jovian satellites. *Icarus* 106: 549–562.
- Schubert G, Spohn T, and Reynolds RT (1986) Thermal histories compositions and internal structures of the moons in the solar

- system. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 224–292. Tucson, AZ: University of Arizona Press.
- Schwarzschild M (1958) *Structure and Evolution of the Stars*, 62 pp. Princeton, NJ: Princeton University Press. (New York: Dover Publications 1965).
- Sears FW and Salinger GL (1975) *Thermodynamics, Kinetic Theory, and Statistical Thermodynamics*, 286 pp. Reading, MA: Addison-Wesley.
- Shakura NI and Sunyaev RA (1973) Black holes in binary systems. Observational appearance. *Astronomy and Astrophysics* 24: 337–355.
- Shapiro SL and Teukolsky SA (1983) *Black Holes, White Dwarfs, and Neutron Stars*, New York: J Wiley Chap. 14.
- Shenk P and Jackson MP (1993) Diapirism on Triton – A record of crustal layering and instability. *Geology* 21: 299–302.
- Shito A, Karato S, and Park J (2004) Frequency dependence of Q in Earth's upper mantle inferred from continuous spectra of body waves. *Geophysical Research Letters* 31: L12603.
- Shock EL and McKinnon WB (1993) Hydrothermal processing of cometary volatiles – Applications to Triton. *Icarus* 106: 464–477.
- Shor VA (1975) The motion of the Martian satellites. *Celestial Mechanics* 12: 61–75.
- Shu F (1984) Waves in planetary rings. In: Greenberg R and Brahic A (eds.) *Planetary Rings*, pp. 513–561. Tucson, AZ: University of Arizona Press.
- Shu F, Johnston D, and Hollenbach D (1993) Photoevaporation of the solar nebula and the formation of the giant planets. *Icarus* 106: 92–101.
- Shultz PH and Lutz-Garihan AB (1982) Forgotten satellites of Mars: A possible record from oblique angle impacts (abstract). *Lunar and Planetary Science XIII*, 698–699.
- Sicardi B, Bellucci A, Gendron E, et al. (2006) Charon's size and an upper limit on its atmosphere from a stellar occultation. *Nature* 439: 52–54.
- Sinclair AT (1978) The orbits of the satellites of Mars. *Vistas in Astronomy* 22: 133–137.
- Singer SF (1968) The origin of the Moon and geophysical consequences. *Geophysical Journal of the Royal Astronomical Society* 15: 205–226.
- Singer SF (1970) Origin of the Moon by capture and its consequences. *EOS Transactions, American Geophysical Union* 51: 637–641.
- Singer SF (1972) Origin of the Moon by tidal capture and some geophysical consequences. *The Moon* 5: 206–209.
- Singer SF and Bandermann LW (1970) Where was the Moon formed? *Science* 170: 438–439.
- Slattery WL, Benz W, and Cameron AGW (1992) Giant impacts on a primitive Uranus. *Icarus* 99: 167–174.
- Smith BA, Soderblom LA, Banfield D, et al. (1989) Voyager 2 at Neptune – Imaging science results. *Science* 246: 1422–1449.
- Smith DE, Lemoine FG, and Zuber MT (1995) Simultaneous estimations of the masses of Mars, Phobos and Deimos from distant spacecraft encounters. *Geophysical Research Letters* 22: 2171–2174.
- Smith JC and Born GH (1976) Secular acceleration of Phobos and Q of Mars. *Icarus* 27: 51–53.
- Sohil F, Spohn T, Breuer D, and Nagel K (2002) Implications from Galileo Observations on the Interior Structure and Chemistry of the Galilean Satellites. *Icarus* 157: 104–119.
- Solomon SC (1986) On the early thermal state of the Moon. In: Hartmann WK, Phillips RJ, and Taylor GJ (eds.) *Origin of the Moon*, pp. 435–452. Houston, TX: Lunar and Planetary Institute.
- Soter S and Harris AW (1977) The equilibrium figures of Phobos and other small bodies. *Icarus* 30: 192–199.
- Spohn T and Schubert G (2003) Oceans in the icy Galilean satellites of Jupiter? *Icarus* 161: 456–467.
- Stephens DC and Noll KS (2006) Detection of six trans-neptunian binaries with NICMOS: A high fraction of binaries in the cold classical disk. *The Astrophysical Journal* 131: 1142–1148.
- Stern SA, Bottke WF, and Levison HF (2003) Regarding the putative eccentricity of Charon's orbit. *The Astrophysical Journal* 125: 902–905.
- Stern SA and McKinnon WB (2000) Triton's surface age and impactor population revisited in light of Kuiper Belt fluxes: Evidence for small Kuiper Belt objects and recent geological activity. *The Astrophysical Journal* 119: 945–952.
- Stern SA, McKinnon WB, and Lunine JI (1997) On the origin of Pluto, Charon, and the Pluto–Charon binary. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 605–663. Tucson, AZ: University of Arizona press.
- Stern SA, Weaver HA, Steffl AJ, et al. (2006) A giant impact origin for Pluto's small moons and satellite multiplicity in the Kuiper belt. *Nature* 439: 946–948.
- Stevenson DJ (1984) The Uranus–Neptune dichotomy: The role of giant impacts. *Lunar and Planetary Science XVI*: 1011–1012.
- Stevenson DJ (2001a) Origin of the Jovian Satellites. Paper presented at the Jupiter Conference, Boulder, CO, 25–30 Jun 2001.
- Stevenson DJ (2001b) Jupiter and its moons. *Science* 294: 71–72.
- Stevenson DJ, Harris AW, and Lunine JI (1986) Origin of satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 39–88. Tucson, AZ: University of Arizona Press.
- Szebehely V and Evans RT (1980) On the capture of the Moon. *Celestial Mechanics* 21: 259–264.
- Szeto AMK (1983) Orbital evolution and origin of the Martian satellites. *Icarus* 55: 133–168.
- Takeda T and Ida S (2001) Angular momentum transfer in a protolunar disk. *The Astrophysical Journal* 560: 514–533.
- Tanaka H, Takeuchi T, and Ward WR (2002) Three-dimensional interactions between a planet and an isothermal disk. I: Corotation and Lindblad torques and planet migration. *The Astrophysical Journal* 565: 1257–1274.
- Tanaka H and Ward WR (2004) Three-dimensional interaction between a planet and an isothermal gaseous disk. II: Eccentricity waves and bending waves. *The Astrophysical Journal* 602: 388–395.
- Taylor SR and Jakes P (1974) The geochemical evolution of the Moon. In: *Proceedings 5th Lunar Science Conference*, pp. 1287–1305, Houston, TX: Lunar Science Institute.
- Thomas PC (1993) Gravity, tides, and topography on small satellites and asteroids – Application to surface features of Martian satellites. *Icarus* 105: 326–344.
- Thompson C and Stevenson DJ (1988) Gravitational instability in two-phase disks and the origin of the Moon. *The Astrophysical Journal* 333: 452–481.
- Tolson RH, Duxbury TC, Born GH, et al. (1978) Viking first encounter of PHOBOS – Preliminary results. *Science* 199: 61–64.
- Touma J and Wisdom J (1993) The chaotic obliquity of Mars. *Science* 259: 1294–1297.
- Touma J and Wisdom J (1994) Evolution of the Earth–Moon system. *The Astrophysical Journal* 108: 1943–1961.
- Touma J and Wisdom J (1998) Resonances in the early evolution of the Earth–Moon system. *The Astrophysical Journal* 115: 1653–1663.
- Tsiganis K, Gomes R, Morbidelli A, and Levison H (2005) Origin of the orbital architecture of the giant planets of the solar system. *Nature* 435: 459–461.
- Veras D and Armitage PJ (2004) Outward migration of extrasolar planets to large orbital radii. *Monthly Notices of the Royal Astronomical Society* 347: 613–624.
- Wada K, Kokubo E and Makino J (2006) High-resolution simulations of a Moon-forming impact and post-impact evolution. *The Astrophysical Journal* 638: 1180–1186.

- Ward WR (1979) Present obliquity oscillations of Mars – Fourth order accuracy in orbital E and I. *Journal of geophysical Research* 84: 237–241.
- Ward WR (1981) Orbital inclination of Iapetus and the rotation of the Laplacian plane. *Icarus* 46: 97–107.
- Ward WR (1993) Density waves in the solar nebula – Planetary velocities. *Icarus* 106: 274–287.
- Ward WR (1997) Protoplanet migration by Nebula tides. *Icarus* 126: 261–181.
- Ward WR (2002) *Astrophysical Tides: The Effects in the Solar and Exoplanetary Systems*. Nanjing, China: IAU colloquium.
- Ward WR and Cameron AGW (1978) Disc evolution within the Roche limit. *Lunar and Planetary Science Conference* 9: 1205–1207.
- Ward WR and Canup RM (2000) Origin of the Moon's orbital inclination from resonant disk interactions. *Nature* 403: 741–743.
- Ward WR and Canup RM (2003) Viscous evolution of an impact generated water/rock disk around Uranus. *Bulletin of American Astronomical Society* 35: 1046.
- Ward WR and Canup RM (2006) Forced resonant migration of Pluto's outer satellites by an impact-produced Charon. *Science* 313: 1107–1109.
- Ward WR and Hahn JM (1994) Damping of orbital inclinations by bending waves. *Icarus* 110: 95–108.
- Ward WR and Hahn JM (2000) Disk–planet interactions and the formation of planetary systems. In: Mannings V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, pp. 1135–1158. Tucson, AZ: University of Arizona Press.
- Weaver HA, Stern SA, Mutchler MJ, et al. (2006) Discovery of two new satellites of Pluto. *Nature* 439: 943–945.
- Weidenschilling SJ (1977) Aerodynamics of solid bodies in the solar nebula. *Monthly Notices of the Royal Astronomical Society* 180: 57–70.
- Weidenschilling SJ (1995) Can gravitation instability form planetesimals? *Icarus* 116: 433–435.
- Wetherill GW (1985) Occurrence of giant impacts during the growth of the terrestrial planets. *Science* 228: 877–879.
- Wetherill GW (1992) An alternative model for the formation of the asteroids. *Icarus* 100: 307–325.
- Wetherill GW and Cox LP (1984) The range of validity of the two body approximation in models of terrestrial planet accumulation. I. Gravitational perturbations. *Icarus* 60: 40–55.
- Wetherill GW and Cox LP (1985) The range of validity of the two body approximation in models of terrestrial planet accumulation. II. Gravitational cross sections and runaway accretion. *Icarus* 63: 290–303.
- Wetherill GW and Stewart GR (1989) Accumulation of a swarm of small planetesimals. *Icarus* 77: 330–357.
- Whipple FL (1964) The history of the solar system. *Proceedings of the National Academy of Sciences* 52: 565–594.
- Widmer R, Masters G, and Gilbert F (1991) Spherically symmetric attenuation within the Earth from normal mode data. *Geophysical Journal International* 104: 541–553.
- Williams JG, Boggs DH, Yoder CF, Ratcliff JT, and Dickey JO (2001) Lunar rotational dissipation in solid body and molten core. *Journal of Geophysical Research* 106: 27933–27968.
- Wilson EH and Atreya SK (2004) Current state of modeling the photochemistry of Titan's mutually dependent atmosphere and ionosphere. *Journal of geophysical Research* 109: E06002.
- Winters RR and Malcuit RJ (1977) The lunar capture hypothesis revisited. *The Moon* 17: 353–358.
- Wisdom J (2004) Spin-orbit secondary resonance dynamics of Enceladus. *The Astrophysical Journal* 128: 484–491.
- Wood JA (1972) Fragments of terra rock in the Apollo 12 soil samples and a structural model of the moon. *Icarus* 16: 462–501.
- Wood JA, Dickey JS, Marvin UB, and Powell BN (1970) Lunar anorthosites. *Science* 167: 602–604.
- Yoder CF (1982) Tidal rigidity of Phobos. *Icarus* 49: 327–346.
- Young EF and Binzel RP (1994) A new determination of radii and limb parameters for Pluto and Charon from mutual event lightcurves. *Icarus* 108: 219–224.
- Yuan D-N, Sjogren WL, Konopliv AS, and Kucinskas AB (2001) Gravity field of Mars: A 75th degree and order model. *Journal of Geophysical Research* 106: 23377–23401.
- Yung YL, Allen M, and Pinto JP (1984) Photo chemistry of the atmosphere of Titan – Comparison between model and observations. *Astrophysical Journal Supplement Series* 55: 465–506.
- Zolotov MY, Owen T, Atreya S, Niemann HB and Shock EL (2005) An endogenic origin of Titan's methane. *American Geophysical Union, Fall Meeting 2005*, abstract #P43B-04.

Relevant Websites

- www.lasp.colorado.edu – Laboratory for Atmospheric and Space Physics.
- www.ssd.jpl.nasa.gov – NASA Jet Propulsion Laboratory, Solar System Dynamics.

10.15 Interiors and Evolution of Icy Satellites

H. Hussmann, Institut für Planetologie, University of Münster, Münster, Germany

C. Sotin, CNRS, Nantes, France

J. I. Lunine, Istituto di Fisica dello Spazio Interplanetario, Rome, Italy, The University of Arizona, Tucson, AZ, USA

© 2007 Elsevier B.V. All rights reserved.

10.15.1	Introduction	509
10.15.2	Spectroscopic Constraints on Composition	510
10.15.3	Elemental Abundance from Density	511
10.15.4	Size, Shape, and Mass	513
10.15.5	Modeling the Interior Structure	514
10.15.5.1	Density and Composition	514
10.15.5.2	Gravity	515
10.15.5.3	Two- and Three-Layer Structural Models	516
10.15.5.4	Pressure-Induced Phase Transitions	518
10.15.5.5	Oceans	518
10.15.5.6	Thermal State	520
10.15.5.7	Heat Transfer	521
10.15.5.8	Effect of Volatiles	522
10.15.6	Evolution of Satellite Interiors	523
10.15.6.1	Tidal Heating	523
10.15.6.2	Heat Balance	524
10.15.6.3	Thermal Histories	525
10.15.7	Interior Structure of Selected Icy Satellites	525
10.15.7.1	Ganymede and Callisto	526
10.15.7.2	Europa	527
10.15.7.3	Titan	530
10.15.7.4	Enceladus	532
10.15.7.5	Rhea and Iapetus	533
10.15.7.6	Triton	534
10.15.8	Future Prospects for Determining Satellite Internal Structure	535
References		535

10.15.1 Introduction

The icy moons of the outer solar system are numerous and diverse in their properties. Although none of these objects approaches the size of Earth, three are larger in diameter than the planet Mercury, and internal pressures for these approach several gigapascals, sufficient to yield phase changes and significant compression in the water-ice component – but not the rock. Therefore, that water ice would be among the major constituents of these bodies means that the interior physics, in terms of the effect of the phase diagram and equation-of-state of water ice, is as complex as for the Earth, where analogously megabar pressures compel

rock to undergo compressive and phase-change effects. The addition of soluble and generally more volatile compounds to the water ice lowers the melting temperature, yields partial melts, and greatly complicates the treatment of the interiors of the icy satellites.

As was already noticed by Lewis (1971), the large icy satellites may contain global liquid layers underneath their icy surfaces. The presence and the extent of such liquid layers will depend on the thermal state and composition of the H₂O layer in the outer part of the satellite. Since solar insolation is only of minor importance for the heat budget of the satellites in the outer solar system, the energy required to form and

maintain liquid layers has to be provided from the interior. Radiogenic heating in the rock component and, in some cases, tidal friction are the main heat sources for satellites. The presence of volatiles such as ammonia, significantly reducing the melting temperature of ice, will be crucial for the existence and maintenance of these internal oceans. Lewis (1971) used an equilibrium condition with the heat production due to radioactive nuclides being equal to the heat loss through the ice shell. However, the assumption of equilibrium may not be valid throughout the satellite's history, especially at the early stages of evolution. Time-dependent thermal histories and the consequences for internal structure were calculated by Consolmagno and Lewis (1976, 1978) for the icy Galilean satellites, and for several size classes of model satellites, respectively. These models, however, suffered from the lack of reliable data, especially for the mean densities deduced from the masses and the radii, of the outer planet satellites.

With the Voyager flybys (1979, Jupiter; 1981, Saturn; 1986, Uranus; 1989, Neptune) data on the physical properties and images of the surfaces of the satellites improved significantly. The new constraints led to refinements of the then available models (Schubert *et al.*, 1981, 1986; Lupo, 1982; Ellsworth and Schubert, 1983), and, due to the observed diversity of surface features of the satellites, to the development of more complex models for individual satellites. By the end of the 1970s, the significance of tidal heating for some of the outer planet satellites became apparent (Peale *et al.*, 1979), leading to considerable interest in thermal and interior structure models of Io, Europa, Ganymede, and Enceladus, the satellites for which tidal heating was considered to be the cause of volcanic activity (Io) and intense resurfacing in the satellites' history.

The Galileo mission (Jupiter orbiter, 1995–2003) led to further improvement of the knowledge on the internal structure of the Galilean satellites (for a post-Galileo review, see Schubert *et al.*, 2004). Until recently, the Galilean satellites were the only satellites in the outer solar system for which higher gravitational moments were determined from space-craft flybys. Potential theory and the theory of equilibrium figures in combination with the acquired data allow for a more detailed inference of the internal structure of satellites if their rotational state and the tidal potentials are known (Dermott, 1979; Zharkov *et al.*, 1985). Galileo data have been used to construct models of the four Galilean satellites

(Anderson *et al.*, 1996, 1998, 2001a, 2001b; Sohl *et al.*, 2002; Schubert *et al.*, 2004), consisting of ice, rock, and metals. The data provide evidence for a considerable degree of internal differentiation for the three inner Galilean satellites. Callisto, however, is only partially differentiated (Anderson *et al.*, 2001b) which may be a consequence of different accretion scenarios (Canup and Ward, 2002; Estrada and Mosqueira, 2006) and orbital evolution of Ganymede into the Laplace resonance, an orbital commensurability between the three inner Galilean satellites, Io, Europa and Ganymede (Showman and Malhotra, 1997; Peale and Lee, 2002).

The acquisition of similar data is expected from the Cassini mission (orbiter in the Saturn system since Jul 2004) for some of the icy satellites of Saturn. At the time of this writing gravity data were already obtained from flybys at Titan and Rhea. However, the analysis has not been completed yet, and further information will be acquired also for other satellites in the mission's course.

In this chapter, we review the composition of icy satellites, describe the construction of interior structure models, present recent models composed of rock (including iron) and ice, describe how other materials affect the interior, and conclude with recent discoveries and key issues. We focus generally on the larger moons (**Figures 1** and **2**), those for which compression is important but porosity is not (**Figure 3**), and will briefly discuss some smaller moons, for example, Enceladus, where recent Cassini data have constrained particular properties of the interior.

10.15.2 Spectroscopic Constraints on Composition

Water ice is the major carrier of oxygen in the solid material of the outer solar system and is abundant in the moons of the giant planets. Spectroscopic observations, mostly in the near-infrared, constrain the presence and amount of water ice in the crust. Callisto's surface contains 20–45% by weight water ice (Calvin and Clark, 1991), while Ganymede's surface ice content is even larger at upwards of 90%, and Europa's approaches 100% (Clark, 1980). Other oxygen-bearing compounds are present as well, such as hydrogen peroxide on Europa and CO₂ on Ganymede (Hibbets *et al.*, 2003). In the Saturn system, the intermediate-sized satellites Mimas, Enceladus, Tethys, Dione, Rhea, and Hyperion show water-ice absorption features on their surface (Cruikshank *et al.*,

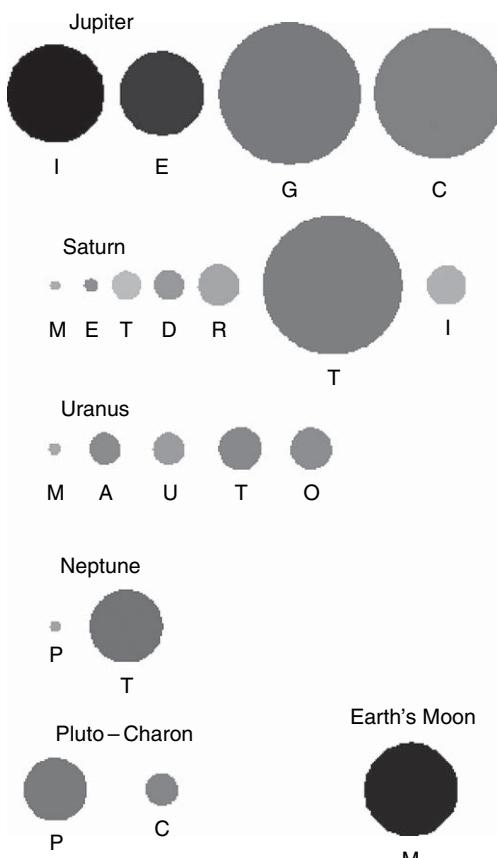


Figure 1 The largest moons of the solar system. Shown are the satellite systems of Jupiter, Saturn, Uranus, Neptune, and the Pluto–Charon system, including all moons with a radius larger than about 200 km. The smallest included here is Mimas with a radius of 198.8 km. Within each system, the satellites are ordered by the distance from the planet (distances are not to scale). The shading illustrates the mean density relative to each other, with Io being the densest (3500 kg m^{-3}) and Tethys the one with the smallest density (956 kg m^{-3}). Pluto and the Moon which is slightly smaller than Io are included for comparison. From left to right (and top to bottom): Io, Europa, Ganymede, Callisto; Mimas, Enceladus, Tethys, Dione, Rhea, Titan, Iapetus; Miranda, Ariel, Umbriel, Titania, Oberon; Proteus, Triton; Pluto, Charon; the Moon.

2005), as does Phoebe (Clark *et al.*, 2005), as the major, but not sole (Brown *et al.*, 2006), spectroscopically active component of their surfaces. Measurement of geyser material coming from the subsurface of Enceladus shows both CH_4 and CO_2 (Waite *et al.*, 2006), and it is thus unclear whether the methane was present at the time of Enceladus' formation or formed from CO_2 (Matson *et al.*, 2007). Titan's surface, beneath its haze-shrouded atmosphere, also contains water ice (Tomasko *et al.*, 2005). In the Uranian

system, water ice has been identified as the major constituents on Miranda (Brown and Clark, 1984), Ariel, and Umbriel, and presumably is on Titania and Oberon as well, but the abundance relative to other constituents is poorly constrained. Water ice has been detected on Neptune's moon Triton, and on Pluto and its moon Charon (see Chapter 10.16).

10.15.3 Elemental Abundance from Density

The major solid-forming materials in the outer solar system, based on simple solar elemental abundance considerations of the relevant elements, are water ice, silicates, and iron compounds. These are the materials which we will call, for brevity's sake, respectively, 'ice' and 'rock'. Where we consider 'ices' other than water ice, for example, solid carbon dioxide or ammonia–water mixtures, we will specifically call these out. Likewise, we will consider the rock phase to explicitly include the corresponding solar abundances of iron and nickel, the primary metals. Only where required (specifically in the case of Ganymede, with its metal core distinct from the silicates) we will consider these separately. To first order, then, our models of icy satellites consist of ice and rock, including the iron-component. Chondritic abundances are assumed for the composition of the rock and iron component.

The significance of the densities of the icy satellites lies in the constraints they impose on interior models, and in turn (for satellites that are subjected to strong compressional effects) these models are the intermediary between the measured densities and the corresponding rock-to-ice ratios within the satellites. Whether water dominates over rock, or vice versa, is a strong function of whether carbon is mostly locked up in oxidized or reduced molecules, because the solar elemental C/O ratio is close to one-half according to the most recent determinations (Table 1; Grevesse *et al.*, 2005). The revised solar abundances reviewed in Grevesse *et al.* (2005) render the ice-to-rock ratio much more sensitive to the molecular state of the carbon than did the previous set it replaced. However, a number of processes can alter the rock-to-ice ratio in outer planet satellites from the value predicted by solar abundance and a particular assumed value of oxidized carbon (which would be mostly carbon monoxide, but including carbon dioxide as well). The process of satellite disk formation may alter molecular abundances and selectively remove either ice or rock (Canup and Ward,

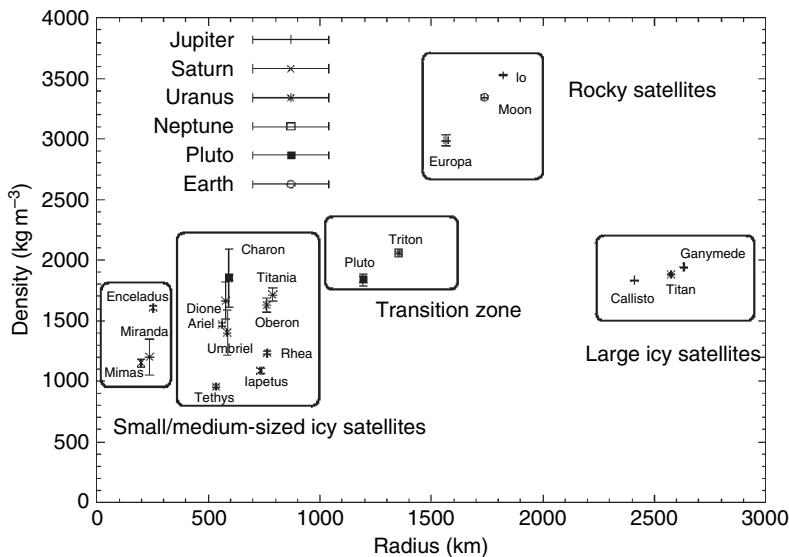


Figure 2 Mean density as a function of radius for all satellites larger than about 200 km and Pluto. Different symbols represent the respective primary planets. Several groups can be distinguished with respect to size and rock content. Adapted from Hussmann H, Sohl F and Spohn T (2006) Subsurface oceans of medium-sized outer planet satellites and large trans-Neptunian objects. *Icarus* 185: 258–273.

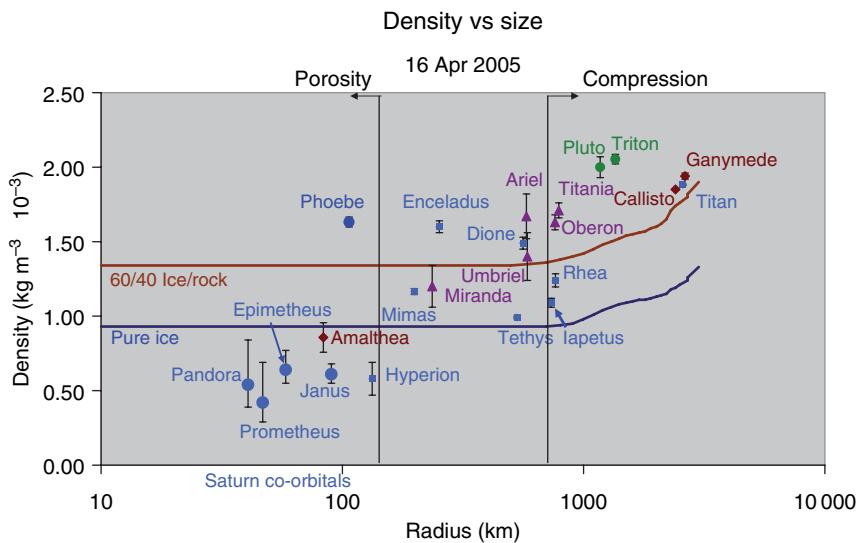


Figure 3 Bulk (measured) density of icy satellites vs satellite radius. Shown also as vertical lines are rough limits on the size above which compressional effects alter the bulk density from the material density of the objects, and below which porosity becomes potentially important in lowering the density. The two approximately horizontal lines show run of model density vs radius for water-ice and silicate bodies containing 60% water ice (top line) and 100% water ice (bottom line). Satellites are color coded according to giant planet system they belong to; Cassini continues to improve the density determinations for the Saturnian satellites. Figure created by Torrence V. Johnson, used with permission.

2006; Estrada and Mosqueira, 2006; see Chapter 10.14). The formation of individual satellites may selectively remove (usually) water, especially in giant collisions or when the satellite has grown sufficiently large that the release of energy per unit mass associated with accretion exceeds the vaporization latent heat per unit

mass of water. If the abundance of organics of high carbon number (and thus mostly refractory) is large enough to remove carbon from the system, then the relationship between water abundance and oxidized carbon species weakens considerably, as shown in Figure 4.

Table 1 Solar photosphere abundances of relevant elements

O	0	± 0.05
C	-0.27	± 0.05
N	-0.88	± 0.06
Si	-1.15	± 0.04
Mg	-1.13	± 0.09
Fe	-1.21	± 0.05

\log_{10} (abundance) relative to oxygen = 0.

Source: Grevesse *et al.* (2005).

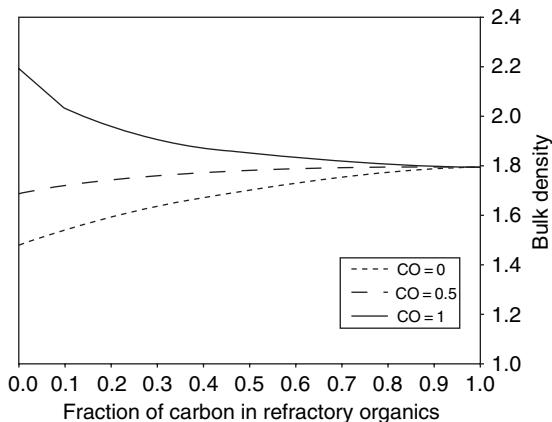


Figure 4 Plot of bulk density of an uncompressed (small) rock-ice moon versus amount of carbon present in refractory organics. Curves for the ratio of carbon monoxide to the total (methane plus carbon monoxide) are shown. For most (>50%) of the carbon locked in refractory (non-CO and non-hydrogen-bearing) organics, the oxidation state of the carbon in the disk from which the satellite formed has little effect on the bulk density. Based on results in Johnson and Lunine (2005).

This last case is particularly interesting because a variety of high-carbon-number organics were likely present in condensed phases in the solar system. Moderate reductions in the oxygen abundance in the inner part of the solar nebula, for example, associated with diffusive redistribution of water vapor and cold trapping of water ice (Stevenson and Lunine, 1988) will result in the gas-phase thermochemical production of acetylene, hydrogen cyanide, and other light organics (Cyr *et al.*, 1999). Chemistry in nebular shocks, photochemistry on the surface of disks, or particle chemistry associated with cosmic rays that penetrate the disk, will produce a variety of organic molecules with higher carbon-to-hydrogen ratio than methane. It is possible that we are seeing evidence of this chemistry in the detection of higher hydrocarbons in comets, and organic phases in Phoebe, likely an object captured by Saturn from solar orbit (Johnson and Lunine, 2005;

Clark *et al.*, 2005). The large amount of carbon dioxide seen in each does not necessarily mean that the latter dominated in abundance over carbon monoxide, but may reflect instead the vast difference in volatility between the two. The effect of the presence of more refractory (carbon-rich) organics is to lower the sensitivity of the bulk density to the ratio of carbon monoxide (which consumes water) to methane.

In sum, while it is clear from spectroscopic data, chemical considerations, and satellite bulk densities that water ice is the dominant solid-forming material in the outer solar system, the relative abundances of other species – particularly carbon bearing – remains uncertain and may be highly variable from object to object and from one satellite system to another. Densities and hence rock-to-ice ratios do not provide a direct one-to-one correspondence with particular carbon compositions and hence formation conditions. Further progress will require detailed spectroscopy of surfaces and the ability to measure in detail, in extraordinary cases such as Enceladus, the chemistry of material emitted from the interiors.

10.15.4 Size, Shape, and Mass

With the exception of Titan, satellite surfaces can be observed directly by spacecraft imaging systems and ground-based telescopes, yielding the physical size of these objects which is required to determine the density. The radius of Titan's solid surface, buried beneath hundreds of kilometers of opaque haze, was first revealed by a radiooccultation performed by Voyager 1 (Lindal *et al.*, 1983). For irregular objects, which include moons as large as Jupiter's Amalthea and Saturn's Phoebe and Hyperion (the last with a long dimension exceeding 200 km), the determination of the total volume is more difficult, and typically requires spacecraft observations at reasonable spatial resolution over a significant fraction of the body (e.g., Saturn's moon Phoebe; Porco *et al.*, 2005).

More subtle is to determine the shapes of moons, which reflect the satellites' internal structure under certain conditions (Hubbard and Anderson, 1978; Dermott, 1979; Zharkov *et al.*, 1985). Moons tend to relax to a triaxial ellipsoidal shape determined by hydrostatic equilibrium of the rotating and tidally deformed body. (This refers to the long-term fluid relaxation of a tidally distorted body. It must be distinguished from the tidal deformations due to periodic forces varying on timescales of the orbital period due to the orbital eccentricity or inclination,

non-negligible obliquity, or because of a nonsynchronous rotational state.) Deviations from hydrostatic equilibrium may occur for instance due to internal activity, incomplete relaxation, or due to a ‘frozen in’ structure which has been adjusted to equilibrium conditions different from the present ones. Measurements by Cassini in the Saturn system reveal at least one case (Iapetus) where the shape of the body is best represented by an oblate spheroid (Thomas *et al.*, 2006), inconsistent with the satellite’s present-day rotation. It has been suggested (Castillo-Rogez *et al.*, 2007) that the observed bulge, which did not adjust to the present equilibrium conditions, is due to a faster rotation rate in the satellite’s past. This may also be the case for Enceladus, albeit on a much smaller level. In general, the shape of a satellite provides important clues to the orbital history of the body and to the rigidity over time of its internal layers.

Density and radial mass distribution inferred from the moment of inertia (MoI) factor (see Chapter 10.02) are the key parameters that constrain interior structure models. The mass of a satellite may be determined from two fundamentally different types of observations (Jacobson, 2004): visual (astrometric) observations of satellite motions from Earth-based telescopes or imaging systems on spacecraft, and tracking via radio signal of the path of a spacecraft as it flies close by a moon and experiences the effect of the gravitational pull on its trajectory. The latter is straightforward in terms of the nature of the effect: the stronger the satellite mass for a given flyby distance, the greater the effect on the spacecraft trajectory (measured in terms of the velocity perturbation). The astrometric observations depend on identifying mutual gravitational interactions among satellites or between satellites and the parent planet. The simplest effect – observing the wobble of a planet around the center of mass of a binary planet – moon system – provided the mass ratio of Pluto’s moon Charon relative to Pluto (Olkin *et al.*, 2003) and the system mass (Tholen and Buie, 1990). For giant planet systems, one looks for orbital commensurabilities that lead to orbital changes sensitive to satellite mass. For example, in the Saturn system, Mimas–Tethys and Enceladus–Dione resonances provide masses of these satellites while small satellites at Lagrange points of larger moons also provide mass information on the latter (see Jacobson (2004) for more details). Not all satellite masses can be well determined in this way, and tracking of spacecraft during close flybys remains the definitive mass

determination technique. Values for the Saturnian satellite system, for which four spacecraft have been used to determine densities, are given in table 4 of Jacobson *et al.* (2006).

10.15.5 Modeling the Interior Structure

Due to the lack of measurements obtained by a seismic network, by deep surface penetration, or by ground-penetrating radar, interior structure models of the icy satellites are based on indirect evidence. The following characteristics are the main clues to investigate the structure of these moons: (1) radius and mass, (2) gravity field, (3) magnetic field, (4) surface temperatures and heat flow, (5) composition of surface and atmosphere, (6) shape, and (7) ongoing activity at the surface. In addition to these features characterizing mainly a satellite’s present state, knowledge on its evolution helps to constrain possible internal states. This includes (1) surface geology and tectonics, which are telltale signs of present or past energetic processes, (2) orbital dynamics, being important for satellites evolving due to tides or in resonances, and (3) the chemical environment during accretion, that is, the composition and physical state of the solar nebula and planetary subnebula. The relevance of the items listed above varies for each individual satellite. However, in the following sections, we will illustrate how the combination of certain aspects can be used to gain a broad picture of the interiors of the icy moons.

10.15.5.1 Density and Composition

The basis of the models is the satellite’s radius R and its density ρ (or equivalently its radius and mass M if the shape is known or if spherical symmetry is assumed). To determine the distribution of mass within the icy satellites requires close spacecraft flybys, sometimes repeatedly, in order to resolve more precisely a body’s gravitational field. The spacecraft will move along a trajectory that is perturbed by the various components of the body’s gravitational field expressed as a spherical harmonic series (Hubbard and Anderson, 1978; Zharkov *et al.*, 1985; Rappaport *et al.*, 2001).

Assuming that the icy moons consist of rock (including iron) and nonporous ice, an assumption which is to first order reasonable on the basis of the compositional abundances of the outer solar nebula, the rock and ice-mass fractions can be inferred from

the mean density. Whereas the ice density is well constrained ($\sim 920 \text{ kg m}^{-3}$), the rock density may vary between about 2500 kg m^{-3} for hydrated rock (Schubert *et al.*, 1986) and 8000 kg m^{-3} for pure iron. A typical value of nonhydrated silicate rock is 3500 kg m^{-3} , which will be used in the following as a reference value whenever we refer to rock. However, for most of the satellites, it is unclear if they are differentiated into several distinct layers or homogeneous. Even for the same material, the density is pressure dependent, and thus depth dependent. In contrast to the larger terrestrial planets, however, compression within the rocky layers can be neglected and the rock and ice densities can be regarded as constant. As shown by Sohl *et al.* (2002), even for the largest satellite, Ganymede, the density variations within each layer are small if self-compression is taken into account. However, the internal pressures can be sufficient to cause phase transitions from ice-I to high-pressure ice phases ice-II, -III, -V, or -VI, the densities of which (up to 1300 kg m^{-3}) significantly exceed the ice-I density (e.g., Schubert *et al.*, 1986; Sotin *et al.*, 1998).

10.15.5.2 Gravity

Besides the possibility to exactly determine the mass and radius of a satellite, close spacecraft flybys can provide measurements of the gravitational moments J_2 and C_{22} . The coefficient J_2 describes the polar oblateness mainly caused by the satellite's rotation and can best be determined with a polar flyby; C_{22} describes the equatorial bulge pointing toward the primary, mainly caused by tidal deformation. It can

best be determined with an equatorial flyby. According to first-order potential theory, these coefficients completely determine the satellite's MoI C about the rotational axis, provided that the satellite is in hydrostatic equilibrium. C is defined as

$$C = \frac{8\pi}{3} \int_0^R \rho(r) r^4 dr \quad [1]$$

where $\rho(r)$ is the run of density with radial distance from the interior. **Table 2** gives basic parameters of selected icy moons, including C/MR^2 . The determination of the dimensionless MoI factor $\text{MoI} = C/MR^2$ yields an additional equation to infer the internal structure of a satellite which can be solved simultaneously with the mass-balance equation. The lower the MoI factor, the greater is the concentration of mass toward the center. For a homogeneous sphere $\text{MoI} = 2/5$. The MoI factors of the Galilean satellites (see **Table 2**) were determined from flybys of the Galileo spacecraft and imply considerable degrees of differentiation for Io (Anderson *et al.*, 2001a), Europa (Anderson *et al.*, 1998), and Ganymede (Anderson *et al.*, 1996). Callisto, however, is regarded as partly differentiated (Anderson *et al.*, 2001b).

Spacecraft flybys can be used to infer if hydrostatic equilibrium is attained. The hydrostatic first-order theory yields $J_2 = (10/3)C_{22}$, which can be verified if J_2 and C_{22} were independently determined from at least two close flybys. A ratio of $10/3$ would be a strong argument in favor of hydrostatic equilibrium, although a combination of nonhydrostatic contributions resulting in the same ratio cannot be ruled out entirely.

Table 2 Characteristics of the largest icy satellites, Io and Pluto

		ρ (kg m^{-3})	Radius (km)	Mass (10^{22} kg)	Eccentricity	P_{rot} (days)	Mass of silicates (10^{22} kg)	% Mass of ice	Power (TW)	MoI factor
J	Io	3528	1822	8.938	0.004	1.77	8.938	0	0.536	0.377
	Europa	2970	1569	4.805	0.010	3.55	4.448	7.4	0.267	0.346
	Ganymede	1940	2634	14.850	0.0015	7.15	10.042	32.4	0.603	0.311
	Callisto	1851	2403	10.759	0.007	16.70	6.903	35.8	0.414	0.355
S	Titan	1881	2575	13.453	0.029	15.95	8.793	34.6	0.528	?
	Rhea	1240	764	0.232	0.001	4.52	0.063	73	0.004	?
U	Titania	1710	790	0.353	0.002	8.71	0.205	42.1	0.012	?
	Triton	2054	1353	2.131	0	R5.88	1.526	28.4	0.092	?
	Pluton	2050	1152	1.313			0.938	28.5	0.056	?
N	Charon	2020	593	0.176	<0.001	6.39	0.124	29.5	0.007	?

The first column gives the planet (J for Jupiter, S for Saturn, U for Uranus, and N for Neptune). This table provides the data for the biggest icy satellites. To assess the radiogenic power at present time, we assume that the density of silicates is 3528 kg m^{-3} , a value equal to Io's density. P_{rot} is the rotational period which is equal to the period of revolution for all the major satellites (synchronous state). Values of the MoI-factor come from the compilation of Schubert *et al.* (2004).

In case of hydrostatic equilibrium, C_{22} is related to the MoI through the secular Love number k_f (e.g., Schubert *et al.*, 2004) depending on the internal density distribution. k_f describes the complete relaxation of the body on long timescales (i.e., the fluid response) adjusting to the given tidal and rotational potentials. For a homogeneous body $k_f = 3/2$, independent from its density. The relation of k_f generally depending on the density distribution within the satellite, to the gravitational coefficient C_{22} can thus be used to constrain the satellite's internal structure. For a synchronously rotating satellite to first order (Schubert *et al.*, 2004)

$$C_{22} = \frac{1}{4} k_f \frac{\omega^2 R^3}{GM} \quad [2]$$

where ω is the angular frequency of the orbital and rotational period, and G is the gravitational constant. For the satellite's axial MoI C , it follows (e.g., Kaula, 1968)

$$\frac{C}{MR^2} = \frac{2}{3} \left[1 - \frac{2}{5} \left(\frac{4 - k_f}{1 + k_f} \right)^{1/2} \right] \quad [3]$$

In hydrostatic equilibrium, the extent of the satellite's internal differentiation – whether the ice has separated from the rock, the rock from the metal, thereby segregating into layers according to the density – can be thus obtained from C_{22} . Using eqn [1] and the value of C , the density distribution within a satellite can be constrained. However, due to the inherent ambiguity of the gravity data, such solutions are not unique (*see* Chapters 10.05 and 10.02).

Albeit not proving it, the ratio of C_{22} and J_2 may provide strong evidence if a moon actually is in hydrostatic equilibrium. Such measurements were successfully performed by the Galileo spacecraft at Io, the only Galilean satellite where the tidal and rotational responses could clearly be separated. Satisfying the $J_2 = 10/3 C_{22}$ constraint, Io is very likely in hydrostatic equilibrium (Anderson *et al.*, 2001a). The results are less conclusive for the other Galilean satellites (Schubert *et al.*, 2004). Nevertheless, the hydrostatic theory is frequently used as a first approach to construct interior structure models of the icy satellites, especially for the large ones.

10.15.5.3 Two- and Three-Layer Structural Models

As an illustration, the two-layer model, consisting of a rock core and an H_2O layer, will be discussed (*see*

Chapter 10.02). Such a model may be a reasonable approach for the medium-sized satellites of Saturn and Uranus, in case of full differentiation. Larger satellites will undergo phase transitions from ice-I to high-pressure ice phases. It should also be noted that even for the medium-sized satellites, for example, for Rhea, the phase transition from ice-I to ice-II can occur in the central region, if the satellite is not fully differentiated in ice and rock (Schubert *et al.*, 1986; Castillo-Rogez, 2006; Anderson and Schubert, 2007). However, in case of full differentiation the pressures in the outer ice layer are sufficiently small for the ice-I phase to remain stable. For simplicity we do not distinguish between the density of solid ice and that of liquid water, possibly forming subsurface oceans in some icy satellites.

From a simple mass-balance equation for a spherical body, the radius R_c and the mass M_c of the core can be directly obtained if the densities are known:

$$R_c = R \left(\frac{\rho_c - \rho_m}{\rho_c - \rho_m} \right)^{1/3}, \quad M_c = 4/3 \pi R_c^3 \rho_c \quad [4]$$

where ρ_c and ρ_m are the densities of the core and mantle (in this case ice-I), respectively. Results for possible core sizes are shown in Figure 5 for the medium-sized satellites of Saturn and Uranus. Strikingly, there is no decrease of rock content with increasing distance from the planet in the Saturnian system. This stands in contrast to the Jovian system, in which the observed density gradient is consistent with a radial temperature distribution in the Jovian subnebula. It is also noteworthy, that with a mean density of only 956 kg m^{-3} , Tethys is lacking a substantial rock-mass fraction, whereas the moons in its vicinity, Enceladus and Dione, contain the highest rock-mass fractions (55%, and 48%, respectively) of the mid-sized Saturnian satellites. The origin of such an irregular distribution is still an open question. In the Uranian system, variations in the rock content are smaller, ranging between 40% (Ariel) and 60% (Titania). An exception is Miranda (23%), which, however, is significantly smaller than the other satellites of Uranus and had a unique origin (Greenberg *et al.*, 1991). The bulk rock-mass fraction is increasing as we go outwards from Saturn to Uranus, a trend which continues if the Neptunian satellite Triton (74%), and the Pluto–Charon system (70% and 63%, respectively) are included.

For the two-layer and more generally for the N -layer model we may additionally use the MoI

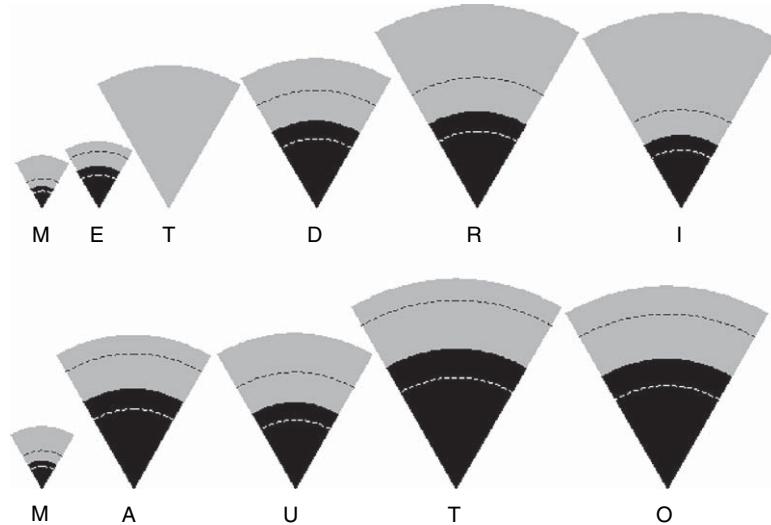


Figure 5 Possible internal structures of the mid-sized icy satellites of Saturn and Uranus consistent with the mean density are shown, assuming complete differentiation into a rock core ($\rho_c = 3500 \text{ kg m}^{-3}$) and an icy mantle ($\rho_m = 960 \text{ kg m}^{-3}$). Complete differentiation is assumed here, only to illustrate the rock content. Whether differentiation was actually completed or whether the satellites consist of homogeneous ice–rock mixtures has to be inferred from spacecraft flybys. For Rhea, for instance, complete differentiation can be ruled out on the basis of Cassini data (Anderson and Schubert, 2007; less *et al.*, 2006). From left to right (and top to bottom): Mimas, Enceladus, Tethys, Dione, Rhea, Iapetus; Miranda, Ariel, Umbriel, Titania and Oberon. Sizes are to scale. The dashed lines show the possible range of core sizes for a variation of the rock density between 2000 and 6000 kg m^{-3} , corresponding to hydrated (larger cores) and iron-rich rock (smaller cores), respectively. Remarkable is the strong variation of rock content in the Saturnian system. Tethys is essentially a homogeneous icy sphere, whereas the rock-mass fraction, and thus core sizes differ considerably for the other satellites. Additionally, there is no obvious correlation of rock content with increasing distance from Saturn. It should be noted that prior to the determination of the higher moments of the gravitational fields, homogeneous ice–rock mixtures cannot be ruled out. The latter would imply ice-II phases in the central regions of the largest of the mid-sized satellites, Titania, Oberon, and Rhea.

(provided it is determined from flybys) to constrain the internal structure. Counting the layers from inside out, the mass-balance equation and the equation for the MoI have to be solved simultaneously:

$$M = 4/3\pi \left[\rho_1 R_1^3 + \sum_{j=2}^N \rho_j (R_j^3 - R_{j-1}^3) \right]$$

$$\text{MoI} = 2/5 \frac{1}{\rho R^5} \left(\rho_1 R_1^5 + \sum_{j=2}^N \rho_j (R_j^5 - R_{j-1}^5) \right) \quad [5]$$

In this set, the density and the radius of each layer are in principle unknown, except for the outermost layer for which the radius coincides with the total radius ($R_N = R$). Thus, only two of the $2N - 1$ unknown quantities can be obtained from eqn [5], implying an inherent ambiguity of multilayered interior structure models. Further assumptions, for example, on the layer densities are thus required to construct the models. In eqn [5] we assumed that the density within each layer is constant, and thus

neglected compressional effects – an assumption justified even for the largest icy satellites (Sohl *et al.*, 2002).

The Jovian satellite Europa is a three-layer-type example, consisting of an iron core, a rocky mantle, and an outer H_2O layer (Anderson *et al.*, 1998; Sohl *et al.*, 2002; Kuskov and Kronrod, 2005). For Europa the MoI was determined from Galileo flybys (see Table 2) using the assumption of hydrostatic equilibrium, and thus the above set of equations can be fully applied. Provided that the densities of the H_2O layer and the rocky layer are known, the set of eqn [5] can be solved, yielding the density and radius of the core, respectively. Depending on its composition and thus density, the core can be as large as about 50% of Europa's total radius. The outer H_2O shell can have an extension of about 80–170 km (Anderson *et al.*, 1998). A typical result for Europa's structure is shown in Figure 6. However, on the basis of the current data, an H_2O shell on top of a uniform mixture of rock and metals, although more unlikely, cannot be ruled out completely (Anderson *et al.*, 1998).



Figure 6 Models of the internal structures of large icy satellites and Europa. Whereas the internal structure of Europa, Ganymede, and Callisto is constrained by the values of the moment of inertia, the latter has yet to be determined by Cassini flybys for Titan. In particular, no constraint on silicate-metal differentiation is yet available for Titan in contrast to Ganymede.

10.15.5.4 Pressure-Induced Phase Transitions

For a given density profile of a spherical satellite, it is straightforward to calculate the pressure $P(r)$ as a function of the radial distance r from the satellite's center solving the hydrostatic equation

$$dP = -g(r)\rho(r)dr \quad [6]$$

where g is the gravitational acceleration. Besides compression effects, which are of minor importance even for the largest icy satellites, the internal pressure is important for the possible occurrence of phase transitions within the ice or rock layers.

The internal structures of the largest icy satellites, Ganymede, Titan, Callisto, Europa, and Triton, are complex because (1) there is evidence at least for the Galilean satellites that they contain iron cores, (2) pressures are sufficiently high for phase transitions of water ice to occur (at least for Ganymede, Callisto, and Titan), and (3) internal heat sources may be sufficient to maintain liquid-water oceans underneath the ice-I shell (Spohn and Schubert, 2003).

Up to a pressure of about 0.2 GPa ice-I will be stable, transforming into ice-III, ice-V, and ice-VI with increasing pressure. In Figure 7, the phase diagram for pure H_2O is shown. A typical interior structure considering the phase transitions and

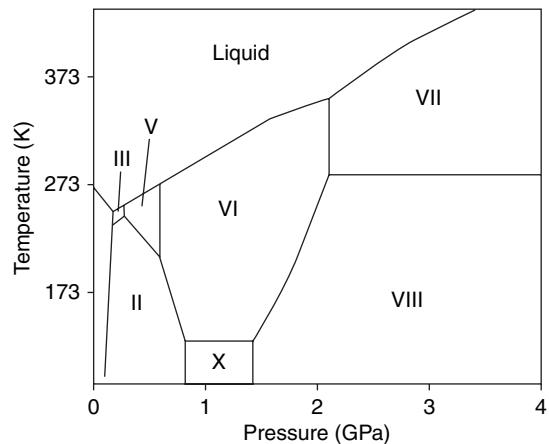


Figure 7 Phase diagram of pure water ice. The different high-pressure ice phases are shown. Most relevant for the icy satellites are ice-I, II, III, V, and VI.

corresponding densities of the high-pressure ice phases is shown as an example in application to Ganymede in Figure 8.

10.15.5.5 Oceans

The fact that the ice-I melting temperature decreases with increasing pressure allows for the existence of liquid layers between the ice-I shell and high-pressure ice phases, if the internal temperature is in the right range (see Figure 9). Volatiles (e.g., ammonia (NH_3)) are important because even small amounts substantially reduce the melting temperature of water ice (Figure 10).

Since the early work of Lewis (1971), the existence of subsurface water oceans in icy satellites and the relevance of ammonia (or other volatiles) were discussed, especially for the largest satellites. Because of possible tidal heating, the Jovian satellite Europa has been investigated in substantial detail. No concurrence with respect to the existence of an ocean was reached in the earlier models (Cassen *et al.*, 1979, 1980; Squyres, 1983b; Ross and Schubert, 1987; Ojakangas and Stevenson, 1989) because of the unknown efficiency of convection in the ice shell. However, the existence of such an ocean is presently considered as the most likely case for Europa. It is consistent with thermal modeling (McKinnon, 1999; Hussmann *et al.*, 2002; Spohn and Schubert, 2003; Tobie *et al.*, 2003, 2005; Mitri and Showman, 2005), with the geological record (Carr *et al.*, 1998; Pappalardo *et al.*, 1998, 1999, and references therein), explains the formation of cycloidal features on the

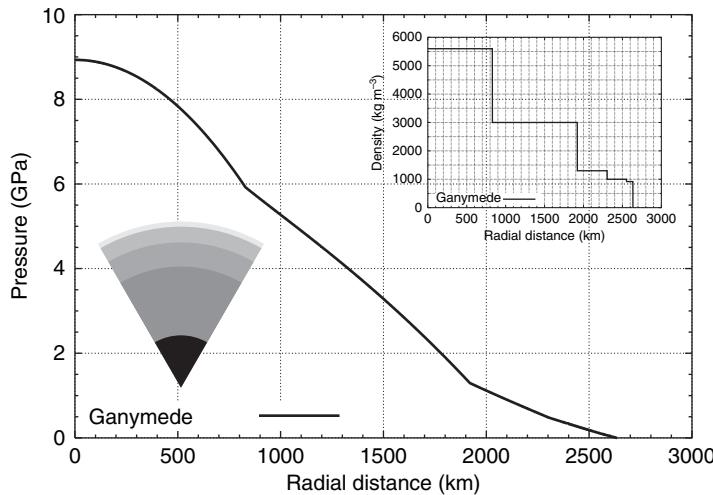


Figure 8 Interior structure model of Ganymede as a type example for a large icy satellite. Shown is the pressure and corresponding density (upper right) as a function of the radial distance from the planetary center. Based on the moment of inertia and mass-balance constraints the internal structure is calculated assuming hydrostatic and thermal equilibrium (equivalent to the Titan models described in Sohl *et al.* (2003)). Further evidence for the differentiation of an iron core comes from the proof of Ganymede’s intrinsic magnetic field during the Galileo mission.

surface (Hoppa *et al.*, 1999) and is consistent with the induced magnetic field detected at Europa (Kivelson *et al.*, 2002).

Induced magnetic fields are generated, because the moons, which revolve around Jupiter in the planet’s equatorial plane, are subject to the time-varying magnetic field of Jupiter. At the location of the moons, the amplitude and direction of the components of the Jovian field vary on a timescale of Jupiter’s rotational period (9.925 h), because of an angle of about 9° between Jupiter’s rotation axis and the axis of the Jovian magnetic field. Due to these variations, an electric current and consequently a secondary (or induced) magnetic field will be generated in electrically conducting regions within the moons. The amplitude and direction of the induced field will be correlated with the variations of Jovian principal field. Such fields were detected at Callisto, Europa, and possibly at Ganymede. For Ganymede, the data analysis is more complicated because Ganymede additionally has an intrinsic magnetic field (Kivelson *et al.*, 2002). The data analysis shows that the induced fields are generated close (a few hundred kilometers, or less) to the surface, ruling out the possibility of currents in the deep interior (e.g., within metallic cores) as a cause for the induced fields (Zimmer *et al.*, 2000). The most probable explanation for the induced fields being generated close to the surface is the electrolytic conductivity due to the presence of salts within a liquid-water layer (Khurana *et al.* 1998;

Neubauer, 1999). The induced magnetic fields are extremely important because it is so far the only observational evidence for the present-day existence of oceans. Surface features, for instance, may have an age of several million years or more and may thus not necessarily represent the ocean’s present state.

Liquid layers can also be expected in other large icy satellites (Lewis, 1971), especially at Titan (Sohl *et al.*, 1995, 2003; Grasset *et al.*, 2000), Ganymede, Callisto (Spohn and Schubert, 2003), and Triton (Lewis, 1971; Ruiz, 2003; Hussmann *et al.*, 2006), provided that there are antifreezes (e.g., ammonia) present. The presence of ammonia can be expected on the basis of models of the solar nebula, condensation temperatures, and theories of satellite formation at least for the Saturn system and beyond (Lewis, 1972). The abundance expected for the icy satellites, based on models for the Saturnian subnebula and on data obtained from the composition of comets and clouds around young stellar objects, is of the order of a few percent (Mousis *et al.*, 2002; Lunine and Stevenson, 1987; Bockelée-Morvan *et al.*, 2004; Dartois and d’Hendecourt, 2001).

For the largest satellites, Ganymede, Titan, and Callisto, internal oceans are expected to be located between the ice-I layer and the high-pressure ice layers. Differentiated Triton and Pluto models imply a pressure regime where the presence of high-pressure phases can neither be ruled out nor confirmed because of the inherent ambiguity of the

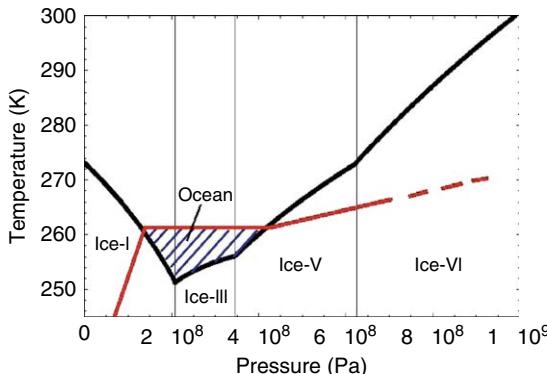


Figure 9 Conceptual sketch of oceans in icy moons. The melting curve of ice-I is shown in black (compare also with **Figure 6**) as a function of internal pressure (and thus with increasing depth). The red curve representing a possible temperature profile intersects the melting curve at two points. In between, the solid ice-phase will not be stable and a liquid layer will form, in this case between the ice-I and ice-V layers. The blue area shows the range where oceans are possible, even for lower temperatures. The higher the temperatures the greater is the vertical extension of the ocean. The minimum temperature allowing for oceans is about 250 K in case of pure water ice. It may be substantially smaller with the presence of ammonia in the ocean (not shown here). The vertical lines mark the locations of the phase transitions.

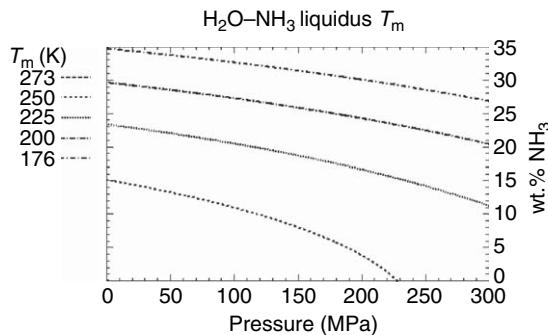


Figure 10 Contours of the melting temperature as a function of ammonia concentration and pressure for an H₂O-NH₃ mixture according to the parametrization of Leliwa-Kopystyński *et al.* (2002). Adapted from Hussmann H, Sohl F, and Spohn T (2006) Subsurface oceans of medium-sized outer planet satellites and large trans-Neptunian objects. *Icarus* 185: 258–273.

models. A structure similar to the one of Europa – with an ocean being in direct contact with the silicate layer – is, however, conceivable for Triton and Pluto.

For the mid-sized satellites, present-day oceans could only exist in the largest members of this class of objects under certain conditions. Besides Ganymede, Titan, Callisto, and Europa, the best candidates for present-day subsurface oceans are Triton,

Pluto, Titania, Oberon and Rhea, provided that they are differentiated and contain ammonia (Hussmann *et al.*, 2006). However recent modeling, based on gravity data from the Cassini mission, suggests that Rhea is not fully differentiated (Anderson and Schubert, 2007; Iess *et al.*, 2006) ruling out the possibility of an internal ocean. Rhea's structure is then likely to be an ice–rock mixture including ice-II in its central region (Castillo-Rogez, 2006). Although it is substantially smaller, Enceladus may also be added to the list of possible ocean candidates because of its strong endogenic and surface activity (Porco *et al.*, 2006; see also Section 10.15.7.4).

The precise determination of the gravity field may provide an indirect way to detect subsurface oceans, due to the greater deformation of the outermost ice-I layer as a response to tidal forces, in case of an ocean. For a moon on an eccentric orbit, the gravitational force exerted by the parent planet varies in a predictable way and it is possible, with flybys at apoapse and periapse (farthest and closest distance to the parent planet) to obtain the time-varying tidal potential, a combination of the external tidal potential (caused by the planet) and the internal or secondary tidal potential caused by the redistribution of mass inside the satellite (i.e., due to the formation of a tidal bulge). The amplitude and phase lag of the satellite's tidal bulge as a response to the periodic external force can be used to determine the overall rigidity of the interior – which in turn constrains the possible presence of interior liquid layers. A liquid layer will decouple the outer ice shell from the deep interior allowing for larger deformation rates of the ice shell due to the tidal forcing as compared to the case where an entirely solid ice shell is directly connected to the rock. Generally, tidal amplitudes will be about an order of magnitude larger in the presence of a liquid layer (Sohl *et al.*, 1995; Moore and Schubert, 2000, 2003). The impact of the liquid layer on the tidal response of Titan's ice shell may provide indirect evidence for a subsurface ocean during the Cassini mission (Castillo *et al.*, 2002). If, and how, the tidal response of the ice shell can be used to infer the thickness of the ice shell and therefore the depth in which liquid water may be present, is currently being investigated for possible future Europa missions (Wahr *et al.*, 2006).

10.15.5.6 Thermal State

The interior structure is linked to a satellite's thermal state via the heat-production rate in the rock

component due to long-lived radiogenic isotopes (see Chapter 10.09). Assuming an elementary abundance of uranium, thorium, and potassium, consistent with the chondritic composition, the heating rate can be obtained from the laboratory-derived half-lives and the energy release per decay of the individual isotopes. The specific present-day heating rate H in the rocky component of the icy satellites is of the order of several 10^{-12} W kg $^{-1}$ which is roughly one order of magnitude less than the heating rate at the very beginning of the solar system. Here, and in the following we neglect short-lived isotopes which may have played a role at the early stages of solar system evolution, especially for planetesimals and their remnants, the asteroids and other small objects (Merk *et al.*, 2002). From a satellite's rock-mass fraction m_{rock} , its overall present-day heat production is given by $H_{\text{total}} = Hm_{\text{rock}}$. Since the heat-transfer rate Q through the ice shell will depend on its thickness and physical state, the latter can be related to the heat-production rate, determined by the chosen density and size of the rock core, if a steady-state $H_{\text{total}} = Q$ is assumed. Although such an equilibrium state may not be exactly matched, it is a reasonable approach for the icy satellites at late stages of their evolution, after primordial heat sources due to accretion, core formation, or early tidal heating have become negligible. However, it should be noted that for the Earth – the only planetary body for which reliable data have been obtained – only 50% of the surface heat flux can be attributed to radiogenic heating. The corresponding values are unknown for other planets and satellites and therefore the equilibrium condition must be regarded as a first assumption. Because of their small size and therefore greater surface-to-volume ratio (in comparison with the Earth and other terrestrial planets), the icy satellites cool faster and approximation to equilibrium is more likely for their present states. Deviations due to additional contributing heat sources, and possible disequilibrium between internal heating and the output of heat through the surface, however, cannot be ruled out completely.

The conductive heat flow through an ice shell of thickness D and temperature T_0 and T_s at the base and the surface of the ice shell, respectively, is given by

$$Q = 4\pi R^2 k \frac{T_0 - T_s}{D} (1 - D/R) \quad [7]$$

assuming a constant value for the thermal conductivity, generally being temperature dependent. Considering a wide temperature range of 40–273 K, k varies for ice-I between 14 and 2 W m $^{-1}$ K $^{-1}$,

according to the parametrization of Klinger (1981), $k(T) = 567/T$, which is valid for temperatures above 30 K. Similar values are obtained using the parametrization of Hobbs (1974). The factor $(1-D/R)$ takes into account the spherical geometry and may be neglected for large satellites. The corresponding temperature profile is given by

$$T(r) = T_s + \frac{T_0 - T_s}{1 - R_0/R} \left(\frac{R_0}{r} - \frac{R_0}{R} \right) \quad [8]$$

with $R_0 = R - D$ the radius of the base of the ice shell. Both equations are connected through Fourier's law. Here, the thermal conductivity and the surface temperature are taken to be constant with the latter ranging from about 100 K for the Jupiter system to about 40 K for Triton and the Pluto–Charon system. The values for individual satellites depend on surface characteristics, for example, albedo, or the presence of an atmosphere (e.g., Titan's surface temperature = 94 K, significantly exceeding the value (about 80 K) expected without an atmosphere). Setting the heat production into equilibrium with the heat flow, $Q = H_{\text{total}}$, the temperature at the base of the ice shell is given by

$$T_0 = T_s + \frac{DH_{\text{total}}}{4\pi R^2 k (1 - D/R)} \quad [9]$$

The above equations are valid for purely conductive shells, only. The heat transfer through a thick and relatively warm ice shell, however, may be dominated by the much more efficient process of subsolidus convection. Whereas at present condition is the main heat transfer process for the small and mid-sized icy satellites, thermal convection can be the dominant process in the large icy satellites.

10.15.5.7 Heat Transfer

Thermal convection is discussed in detail in Chapter 10.09. Whereas the basic equations describing convection in an ice shell are the same as for mantle convection, the physical properties, especially the temperature range in which ices are subject to viscous flow, are completely different as compared to rock. The viscosity, which is difficult to describe by first principles or by parametrizations based on laboratory measurements, is a key parameter controlling convection. It is strongly temperature dependent, and different parameters have to be applied for various creep mechanisms operating in specific temperature and stress regimes. Furthermore, the viscosity is pressure and grain-size dependent. The physical properties of water

ice, creep mechanisms, and laboratory data with regard to the dynamics of icy satellites are discussed, for example, in Weertman (1983), Durham and Stern (2001), Goldsby and Kohlstedt (2001), Sotin *et al.* (1998), and McKinnon (2006).

Due to the large viscosity contrast between the base of the ice shell and the surface, the icy satellites operate in the stagnant-lid regime (Solomatov, 1995; see Chapter 10.09), characterized by a convective region and a relatively thick lid, in which the viscosity is too large for flows to occur. In the lid, the heat transfer is purely conductive and can be calculated by the equations given above. A schematic temperature profile is shown in Figure 11. The convective region is characterized by a small temperature drop, corresponding to a viscosity contrast of about an order of magnitude only. Since pressure changes are small

within the ice shell, adiabatic temperature changes can be neglected and except for the boundary layers, the temperature T in the well-mixed convective region can be regarded as constant. Several scaling laws, based on laboratory data and numerical experiments, have been proposed to describe both the heat flow and the onset of convection (Davaille and Jaupart, 1993; Solomatov, 1995; Grasset and Parmentier, 1998; Deschamps and Sotin, 2001). Non-Newtonian viscosities and the dependence on grain sizes for strongly temperature-dependent convection have been discussed by Barr *et al.* (2004) and McKinnon (2006).

In contrast to previous models (e.g., Schubert *et al.*, 1979), the upper bound of the convective region is not determined by a fixed isotherm, but is found to be dependent on the temperature within the convective region itself (Davaille and Jaupart, 1993; Grasset and Parmentier, 1998). Using such an approach, the temperature at the top of the convective layer can be calculated as a function of the internal temperature, determining also the corresponding heat flow out of the convective region if a viscosity law and an appropriate Rayleigh number are defined. In a steady state, the heat flow out of the convective region will be equal to the purely conductive heat flow through the stagnant lid, which can be used to determine the thicknesses of the individual layers. The more vigorous the convection the thicker will be the convective region and the thinner the stagnant lid (for fixed temperatures at the surface and the base of the ice shell). At present convection plays a major role in the ice shells of the largest satellites, for example, ganimede, Titan, Callists, and Europa.

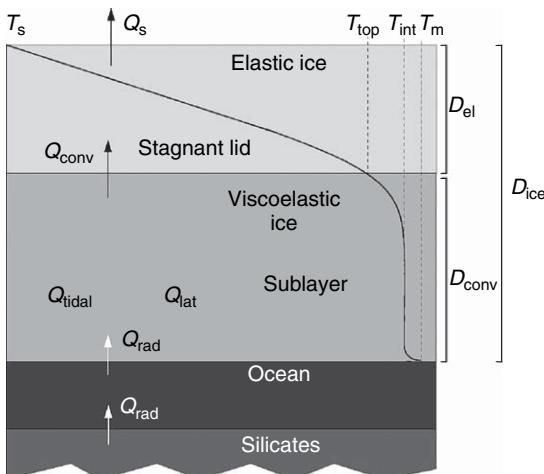


Figure 11 Schematic sketch of a convective ice-I layer at the surface of an icy satellite. Q_{rad} is the radiogenic heating rate from the silicate mantle. Q_{conv} is the heat flow out of the convective region, and Q_s the surface heat flow. Q_{lat} and Q_{tidal} may be additional heat sources in the ice if latent heat is released due to crystallization or in cases where tidal dissipation is an important factor. The ice-I layer is divided into the convective sublayer and the stagnant lid. The temperature (black curve) is high and, apart from the boundary layers at the very top and bottom, almost constant within the convective region. In the stagnant lid, the temperature is decreasing strongly corresponding to the conductive heat transport. The temperature profile can be constructed if the melting temperature T_m , that is, the temperature at the base of the ice shell and the surface temperature T_s , as well as material parameters, for example, the thermal conductivity, are known. Such a structure (not to scale) with a rocky mantle and an H_2O layer on top of it would be appropriate for the satellite Europa. Adapted from Hussmann H, Spohn T, and Wieczorkowski K (2002) Thermal equilibrium states of Europa's ice shell: Implications for internal ocean thickness and surface heat flow. *Icarus* 156: 143–151.

10.15.5.8 Effect of Volatiles

Independent of the mode of heat transport (convection or conduction), one has to check if a solid ice shell could be stable under the given circumstances, that is, whether T_0 is above or below the melting temperature of ice-I. The determination of the pressure P_0 at the base of the shell is straightforward since the mass of the latter is known for a specified thickness D with the remaining mass being included in the sphere below the ice shell. The melting temperature $T_m(P_0)$ can then be determined from the laboratory-derived parametrizations for the given pressure. The reduction of the melting temperature due to a specified small amount of ammonia may be taken into account at this stage (see Figure 10). Parametrizations are given by Chizhov (1993) for pure H_2O , and for $\text{H}_2\text{O}-\text{NH}_3$ mixtures by Hogenboom *et al.* (1997), Leliwa-Kopystynski *et al.*

(2002), and Grasset and Pargamin (2005). If T_0 exceeds the melting temperature the solid ice-I phase will not be stable and a liquid-water layer (ocean) will form underneath the ice shell (**Figure 9**). The upper bound of the ocean will be determined by the melting curve and can be calculated from the corresponding pressure. If T_0 is less than the melting temperature, the ice-I layer will be stable. In this case, one must check to see if the pressure of the ice-I/ice-III or ice-I/ice-II phase transition is reached. If not, the ice shell will consist of ice-I only, and the remaining bulk composition of the satellite will be rock, including the iron-rich components. If the pressure is sufficiently high for the phase change to occur, the lower part of the H₂O layer will consist of high-pressure ice covering the rocky interior.

As mentioned above, a liquid layer may form between the ice-I shell and the high-pressure phase. Such an ocean is assumed to be isothermal because the adiabatic temperature increase with increasing pressure is small for the icy satellites. Therefore, in a simple approach, the temperature at the base of the ocean can be equated with the temperature at the base of the ice shell and the corresponding pressure (and hence depth) at the base of the ocean can again be calculated from the melting curve, this time for the high-pressure phase. Assuming a mean density of the high-pressure ice phase and the rock core, the extension of the latter is determined from the mass-balance constraints. By varying the core size, and thus the heat production in the satellite's deep interior, an equilibrium model can be derived by an iterative process.

Depending on the number of available constraints there may be some ambiguity inherent in the model results and equilibrium may be attained for various core sizes, densities, and ice thicknesses. In **Figure 8**, the interior structure of Ganymede is shown as a typical example of a class of models consistent with the mass balance and MoI constraints assuming thermal equilibrium. Further constraints, for example, the presence of a magnetic field, can be used to restrict possible values of core sizes and densities (e.g., assuming the density of the Fe–FeS eutectic for the core).

10.15.6 Evolution of Satellite Interiors

The presence of subsurface oceans could have been an important factor in the satellites' histories, even for the smaller satellites, because of the greater radiogenic heating rate in the past. Thermal evolution

models can be calculated from heat-balance equations applying the same fundamental principles as for the terrestrial planets (*see* Chapter 10.09). The major modification in application to the icy satellites is the replacement of the physical properties of mantle rock by those of ice. Further complications arise in models in which subsurface oceans play a role. Additional factors, for example, the release of latent heat due to crystallization of an ice layer and the increasing ammonia concentration in the subsurface ocean due to freezing of pure ice have to be taken into account.

10.15.6.1 Tidal Heating

Besides radiogenic heating additional heat sources are relevant especially at the beginning of satellite evolution. The most important ones are accretional heating, release of potential energy due to differentiation, and tidal heating (e.g., Schubert *et al.*, 1986). The latter has been very efficient during the phase in which the satellites were not yet in a synchronous rotation state. This state was obtained later in the evolution by all large and mid-sized icy satellites. For synchronously rotating satellites, tidal heating will be most effective (see eqn [10]) for (1) large satellites on (2) eccentric orbits (3) close to the primary with (4) high internal temperatures. These conditions are best fulfilled by the Jovian satellite Io, for which tidal friction still is the dominating heat source (*see* Chapter 10.09). At present, tidal heating significantly contributes to the heat budget of Europa, Enceladus, and possibly Titan. The example of the vigorous activity of Enceladus shows that the internal composition and the thermal state of a satellite could be crucial for tidal effects. With respect to size, distance from the planet, and orbital eccentricity we would expect tidal heating of Mimas instead of Enceladus. That Enceladus, and not Mimas, is internally active may be due to the higher rock content, and thus larger radiogenic heating, or due to different thermal-orbital histories of the two satellites (see Section 10.15.7.4).

Tidal heating may additionally have played a significant role in the early stages of evolution for several other moons involving highly eccentric orbits or nonsynchronous rotation. Ganymede and Triton, both showing a great diversity of surface terrain and tectonic features, are prominent examples for a specific orbital history, most likely involving strong tidal heating (Showman and Malhotra, 1997; McKinnon and Leith, 1995).

Even for synchronously rotating satellites, tidal friction plays a special role for the histories of some moons, because it provides a link between the orbital evolution (because tidal friction in the satellite implies a loss in orbital energy and reduces the eccentricity) and the thermal state of a satellite (internal heating because of friction within the satellite). For a satellite in synchronous rotation with negligible inclination and obliquity, the tidal heating rate is given by (Segatz *et al.*, 1988)

$$H_{\text{tidal}} = -\frac{21}{2} \frac{R^5 n^5 e^2}{G} \text{Im}(k_2) \quad [10]$$

where n and e are the mean motion and orbital eccentricity of the satellite, respectively. $\text{Im}(k_2)$ is the imaginary part of the second-degree potential Love number which depends on the structure (density profile), rheology (e.g., temperature-dependent viscosity of the planetary material), and frequency of the forcing (i.e., the mean motion n). Here, k_2 describes the viscoelastic response of the satellite to the external periodic forcing on the timescale of the orbital period. It has to be distinguished from the previously introduced fluid Love number, which describes a complete relaxed state of a planetary body on long timescales.

The linkage between tidal heating and thermal evolution has been intensely investigated in application to the Galilean satellites because of Io's thermal activity and the resonance locking of the inner three satellites (Peale *et al.*, 1979; Yoder, 1979; Yoder and Peale, 1981; Greenberg, 1982; Squyres *et al.*, 1983a; Peale, 1986; Ojakangas and Stevenson, 1986; Fischer and Spohn, 1990; Spohn, 1997; see Chapter 10.09).

With respect to the icy satellites, tidal heating has been considered mainly in application to Europa to infer the thermal state and thickness of the ice shell (Cassen *et al.*, 1979, 1980; Ross and Schubert, 1987; Ojakangas and Stevenson, 1989; Hussmann *et al.*, 2002; Tobie *et al.*, 2003; Miti and Showman, 2005; see Section 10.15.7.2) and in application to Enceladus to explain the tectonic surface features (Squyres *et al.*, 1983a; Poirier *et al.*, 1983; Ross and Schubert, 1989; Peale, 2003).

Thermal-orbital models have also been applied to explain Ganymede's tectonic surface features by an evolution into the Laplace resonance (Showman and Malhotra, 1997) and to investigate the thermal–orbital history of Europa and its ocean over long timescales (Hussmann and Spohn, 2004). The relation between the orbital history and tidal dissipation in Titan has been discussed by Sohl *et al.* (1995), and by Tobie *et al.* (2005, 2006).

10.15.6.2 Heat Balance

The evolutionary paths of the icy moons are as diverse as is the appearance of their surfaces and their bulk structure. It is therefore impossible to describe all aspects by a simple set of governing equations. However, the basis of all evolution models is an energy balance between the internal heat sources and the heat transport equations depending on viscosity (in case of convection) and material parameters. The distribution of heat sources will depend on the state of differentiation and, in case of tidal heating on the orbital state of the satellite. The initial conditions will furthermore depend on accretional processes and the thermal environment due to solar insolation. The heat-balance equation for a spherical shell with density ρ , heat capacity c , temperature T , and volume V , bounded by the surfaces S_1 and S_2 ($S_1 > S_2$) is given by (Schubert *et al.*, 1986)

$$\int_V \left(\frac{d}{dt} (\rho c T) - H_V \right) dV = \int_{S_2} q_2 dS - \int_{S_1} q_1 dS \quad [11]$$

H_V is the total heat-production rate per unit volume, q_1 and q_2 are the heat fluxes (power per unit surface) into the shell (through surface S_1), and out of the shell (through surface S_2), respectively. Appropriate equations describing the convective or conductive heat flows have to be applied to calculate q_1 and q_2 . Both parametrized convection models and numerical approaches solving directly the Navier–Stokes equations for the convective ice shell for certain boundary conditions have been applied to investigate the evolution of icy satellites. Recently, two-dimensional (2-D) numerical models were extensively developed in application to the ice shells of the Galilean satellites (Showman and Han, 2004; Barr *et al.*, 2004). The disadvantage of such an approach, in which fewer parameters are required, is an immense increase in computational time (especially for 3-D models) and the greater effort which is required to couple such models with other theories, for example, tidal heating and orbital evolution. Additionally, in some specific problems where the information on flow patterns as a result of numerical models is not needed, parametrized models may be preferred.

An integration of eqn [10] can be done for several shells with different material parameters or heat sources, describing the individual layers of the satellite. The initial temperature profile required for the integration is usually given by an assumed accretional profile of the satellite (e.g., Schubert *et al.*, 1986).

10.15.6.3 Thermal Histories

Thermal histories have been calculated mainly in application to the Galilean satellites and the satellites of Saturn (Consolmagno and Lewis, 1976, 1978; Schubert *et al.*, 1981, 1986; Lupo, 1982; Ellsworth and Schubert, 1983; Multhaup and Spohn, 2007). Typical evolution scenarios of homogeneous mid-sized icy satellites can be divided into four phases (Ellsworth and Schubert, 1983; Schubert *et al.*, 1986; Multhaup and Spohn, 2007). (1) The initial accretional state is characterized by a cold interior, a temperature increase with radial distance from the center, and a steep thermal gradient at the very surface. (2) The satellite is then heated from within due to radioactive decay of long-lived isotopes. The internal temperature increases and, after a typical time of the order of several tens of million years, the temperature profile is reversed, with the central region being warmer than the outer part of the satellite. (3) Depending on the heat sources available and the size of the satellite, convection may start while the satellite heats up. In this phase, maximum temperatures will be reached, with the satellite being in its most dynamical phase, characterized by a maximum surface heat flow and internal activity. Convection leads to more efficient cooling of the satellite whose temperature then again decreases with time for several gigayears. (4) At a certain time, convection will stop due to the generally decreasing radiogenic heating rate, and the satellite will further cool conductively up to 4.5 Gy, that is, the present time.

According to such a scenario, all the small- and mid-sized satellites are expected to be in the fourth, relatively cold phase, cooling conductively. However, further modifications of that scenario can arise if the internal temperatures exceed the melting temperature of ice-I or ice-I/volatile mixtures within the satellite. In that case, liquid layers may form and the satellite will differentiate, forming a rock core and an outer ice layer, possibly contaminated with rocks. A liquid layer may persist enclosed between the rock and ice layers. If these layers can survive up to the present time will depend on the rock content (and thus heating rate of the satellite) and on the amount of volatiles included in the liquid. Because of freezing of pure ice at the top of the liquid layer, the volatile content in the liquid will increase, thereby impeding crystallization until the eutectic composition is reached. Further cooling will lead to the crystallization of the eutectic mixture. Lewis (1971) already noted the significance of ammonia, as the most important antifreeze component for

icy satellites and concluded that the largest objects still may contain liquid layers in their interiors, a result which is consistent with recent models of icy satellites (Spohn and Schubert, 2003; Hüssmann *et al.*, 2006).

It should be noted that in these evolution models additional heat sources, like tidal heating, which may have been important for individual satellites are neglected. The enormous internal heating and activity of Enceladus (Porco *et al.*, 2006; see also the Section 10.15.7.4) shows that individual satellite histories may be completely different from the general scenario described here. However, most of the mid-sized satellites' surfaces are heavily cratered, showing only little or no signs of recent activity, which is consistent with the general concept of evolution of this class of objects.

For the large icy satellites, more complex evolution scenarios, including core formation or the segregation of rocks within the ice shells, have been applied to understand their physical characteristics. This includes models of the incomplete differentiation of Callisto to explain the substantial differences between Ganymede and Callisto (Kirk and Stevenson, 1987; Mueller and McKinnon, 1988; Nagel *et al.*, 2004; McKinnon, 2006), or the differentiation and evolution of Titan (Grasset and Sotin, 1996; Grasset *et al.*, 2000). The three large icy satellites, Ganymede, Callisto and Titan, are further discussed in the following section.

Besides the physical evolution described so far, the chemical history of the icy satellites, which is directly linked to the composition (including also minor components) of the nebula in which the satellites accreted, plays an important role (e.g., Lunine and Stevenson, 1982; Kuskov and Kronrod, 2005). Models have been developed mainly in application to Titan to understand the unique coupling between interior, surface activity, and atmosphere of this icy moon (Lunine and Stevenson, 1987; Tobie *et al.*, 2006; see also Section 10.15.7.3).

10.15.7 Interior Structure of Selected Icy Satellites

While the densities of the Galilean satellites and Titan were determined to high accuracy during Voyager flybys, it was the Galileo spacecraft that obtained most of the information on the interior structure of the Galilean satellites as it orbited around Jupiter between 1995 and 2003, and Cassini should do the same during a set of dedicated 'radio science' flybys of Titan from 2006 through 2008 (Rappaport *et al.*, 1997). In the following sections,

we discuss a selection of the icy satellites in detail, focusing on results from the Galileo and Cassini missions (except for Triton) with respect to internal structure.

10.15.7.1 Ganymede and Callisto

Ganymede is the largest satellite in the solar system with a radius of 2634 km. Its density of 1940 kg m^{-3} lies in between the density of silicates and that of water. The observation that Ganymede has its own intrinsic magnetic field and forms a small magnetosphere inside Jupiter's vast magnetosphere was a major discovery of the Galileo mission. Ganymede has a clear signature of a permanent dipole (Kivelson *et al.*, 1998). As pointed out by Stevenson (2003), a permanent magnetism explanation is conceivable (Crary and Bagenal, 1998) but unlikely and the most reasonable interpretation is a dynamo in the metallic core (Schubert *et al.*, 1996). More recent investigation (Kivelson *et al.*, 2002) suggests a dipole moment equal to $1.32 \times 10^{20} \text{ A m}^2$. The dipole is tilted by 176° , almost antiparallel to the spin axis. The few flybys allow only for the determination of the degree 1 and 2 coefficients of the spherical harmonic decomposition of the magnetic field. By determining the loss of power between degree 1 and 2, the radius of a liquid iron core has been evaluated around 200 km (Kivelson *et al.*, 2002). By comparison with Earth, a liquid Fe–S core is expected although the nature of the light element is still debated for the Earth. Because of the self-sustained magnetic field we can assume that within Ganymede the metal-rich component has separated from the rock forming an iron core, surrounded by a silicate mantle. The outer layers consist of H₂O, which will be present in the form of high-pressure ice, ice-I, or liquid water, depending on internal pressure, temperature, and composition as a function of depth.

A dipolar field cannot fully explain the magnetic data. Two models have been proposed by Kivelson *et al.* (2002) in order to explain the remaining signal: either quadrupole terms resulting from the dynamics of the core dynamo or an induced magnetic dipole driven by the time-varying component of the externally imposed magnetic field of Jupiter's magnetosphere. Both models give equally satisfactory fits to the data although the inductive field is favored (Kivelson *et al.*, 2002) due to its presence within Europa and Callisto (see below).

The gravity data alone can provide the MoI assuming hydrostatic equilibrium. A value of 0.311 has been

obtained (Anderson *et al.*, 1996), implying that Ganymede is strongly differentiated (Figures 6 and 8). A two-layer model (silicate core and H₂O mantle) cannot explain this very small value if reasonable values are taken for the densities of silicate and H₂O (Anderson *et al.*, 1996). An iron-rich core which is required to explain this small value of the MoI is in complete agreement with the interpretation of the magnetic data. However, the radius of each interface cannot be precisely determined since it depends on the choice made for the density (see Chapter 10.02 on the inherent ambiguity of such models). The thickness of the H₂O layer is in the range of 600–900 km (Deschamps and Sotin, 2001; Sohl *et al.*, 2002). Due to the high pressure value reached at the bottom of the H₂O layer, high-pressure ice phases must exist below depth of roughly 150 km. The H₂O layer is subdivided into ice-Ih, ice-III, ice-V, and ice-VI sublayers (Figure 8).

The state of the H₂O outer layer is a subject of debate. If a liquid layer exists, as suggested by the induced magnetic field (Kivelson *et al.*, 2002), it should be located between the ice-I and the ice-V layer (Figures 6 and 8). If the liquid layer is pure water, and if there is no other internal heating source than the radiogenic decay in the silicates (0.6 TW, Table 2), then the scaling laws describing stationary convection suggest that the ice viscosity should have a value around 10^{15} Pa s (Figure 12). Such a value is much larger than the viscosity of ice measured on terrestrial glaciers. If the melting-point viscosity is comparable with that of terrestrial glaciers ($5 \times 10^{13} \text{ Pa s}$) then the amount of heat that can be removed is on the order of 1.5 TW. Another source of heat is required to explain

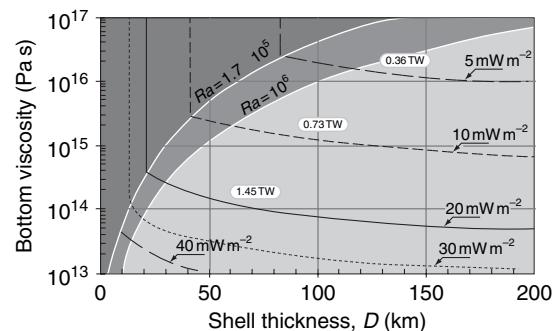


Figure 12 Amount of heat transferred by conduction or by convection within the ice-I shell of a large icy satellite, plotted as bottom viscosity, required to match a given heat flux, vs thickness of the shell. Lines of constant total power and surface heat flow (power per unit area) are shown. Parameters for Callisto have been used.

such a large value. One can think of tidal heating. According to simulations (Showman and Malhotra, 1997; Sotin and Tobie, 2004; Tobie *et al.*, 2005), present-day tidal heating cannot be larger than radiogenic heating and cannot explain such a large value. Another explanation invokes the presence of anti-freeze materials that would lower the temperature (increase the viscosity) at the solid–liquid interface. But this is clearly not proven.

The differentiation of the metallic (largely iron, based on solar abundances) core may have been a late episode in the evolution of the icy satellites (Grasset and Sotin, 1996; Deschamps and Sotin, 2001). This hypothesis presumes that the outer ice layer acts as an insulator that cannot release all the heat produced by the decay of the radiogenic elements. How the internal dynamics and evolution scenarios, including core formation, are related to Ganymede's geology, characterized by two types of terrain – the heavily cratered dark terrain and the less cratered bright terrain – still is an open question. Images obtained by Voyager and Galileo suggest that Ganymede experienced some resurfacing events about 3.5 Ga.

Callisto's radius is about 200 km smaller than that of Ganymede and its mass is 70% that of Ganymede. The old, heavily cratered, and dirt-rich surface suggests that endogenic resurfacing has never happened since it completed its accretion. The post-Galileo view of Callisto is surprising. If hydrostatic equilibrium is achieved, the Galileo gravity data suggest that the MoI is equal to 0.355. This value is not compatible with a fully differentiated Callisto (Anderson *et al.*, 2001b). The magnetic data suggest that an ocean is present at depth around 150 km (Khurana *et al.*, 1998; Zimmer *et al.*, 2000). These two interpretations of geophysical data seem contradictory since the presence of an ocean would lead to differentiation. In order to reconcile these two observations, Nagel *et al.* (2004) propose that Callisto's core is a mixture of ice and rock with the rock concentration near the close-packing limit.

Models of heat transfer in the ice-I outer layer of Callisto vary in their results. One model assumes that convection is not possible, which explains the presence of the ocean (Ruiz, 2001). A more recent paper (McKinnon, 2006) suggests that the cooling from above of the ice-I shell leads to subsolidus convection. This latter work confirms the results of previous modeling studies (e.g., Deschamps and Sotin, 2001). The amount of heat that can be removed by subsolidus convection is the same whether one uses the scaling laws proposed by Dumoulin *et al.* (1999) or

those described in Deschamps and Sotin (2001). The amount of heat transferred depends strongly on the viscosity and weakly on the thickness (Figure 12). This amount is much larger than the heat released by the decay of the radiogenic elements at present time (0.4 TW, Table 2) unless the viscosity of ice near its melting point is larger than 10^{15} Pa s. Convection is not only able to remove the radiogenic heat but also to freeze the ocean very quickly unless some antifreeze contaminants like ammonia are present (Deschamps and Sotin, 2001; McKinnon, 2006).

The comparison between Ganymede and Callisto is interesting since these two satellites have mass and radius that are fairly similar but their internal structure is quite different. Ganymede is fully differentiated with a liquid iron core generating a magnetic dynamo, whereas Callisto is only partially differentiated. Another difference is their surfaces. Ganymede's surface was tectonically active about 3 Ga, whereas Callisto's surface does not show any sign of late activity. More information about their interiors is needed to understand why they are so different. The lower radiogenic power within Callisto may be one explanation. Alternatively, different accretion scenarios, due to different locations in the Jovian subnebula (Canup and Ward, 2002; see Chapter 10.14) may be considered. This would be closely connected with the evolution of Ganymede into the Laplace resonance locking. Both, a primordial origin (Peale and Lee, 2002), or an evolution into resonance at a later stage (Showman and Malhotra, 1997), may have caused a difference between the two satellites with respect to internal structure and geology. The example of Ganymede and Callisto shows that not only the bulk composition but also the interaction with other satellites or with the subnebula during the phase of accretion determines the satellite evolution. Evolution scenarios of icy moons in many cases cannot be considered separated from each other but must be seen in the context of the whole satellite system. The detailed investigation of Titan, which ranges between Ganymede and Callisto with respect to size and mean density, in the course of the Cassini mission may additionally help to understand the extreme diversity of the evolution and present state of this class of large icy satellites (see Figure 2), consisting of Ganymede, Titan, and Callisto.

10.15.7.2 Europa

With respect to the distance from Jupiter, Europa is the second of the Galilean moons (note that there are

additionally four small satellites, Metis, Adrastea, Amalthea, and Thebe, inside Io's orbit, which is the innermost of the Galilean satellites), orbiting around the giant planet at a period of 3.55 days. Its density of 3018 kg m^{-3} lies between that of Io (3530 kg m^{-3}) and those of Ganymede (1936 kg m^{-3}) and Callisto (1851 kg m^{-3}). As described in the previous section, the mass fraction of ice on Europa is only about 6–9% based on the mean density. The potential of the gravity field has been measured during several Europa flybys in the course of the Galileo mission (Anderson *et al.*, 1998; Schubert *et al.*, 2004). Unfortunately, the flybys were not with a geometry that allowed J_2 and C_{22} to be determined independently from each other. Assuming hydrostatic equilibrium, ($C_{22} = 0.3J_2$), it is possible to determine the value of the hydrostatic Love number k_f which is related to the MoI C about the rotation axis by the Radau formula (eqn [3]). The value found for Europa suggests that it is fully differentiated including an iron-rich core, a silicate mantle, and an H_2O shell (Anderson *et al.*, 1998; Sohl *et al.*, 2002; Schubert *et al.*, 2004). Another way to interpret the density data, which cannot be completely ruled out on the basis of the MoI value, is that Europa would be made of a rocky core of 1450 km in radius overlaid by a 110-km-thick layer of H_2O .

The high-resolution images obtained during the Galileo mission provide some clues about Europa's internal structure and evolution. Compared to the other icy Galilean satellites, Europa's surface displays intriguing tectonic features, such as domes, disrupted blocks, chaotic terrains, and fractures (Carr *et al.*, 1998; Greeley *et al.*, 1998; Pappalardo *et al.*, 1998, 1999; Spaun *et al.*, 1998; Greenberg *et al.*, 1998; Riley *et al.*, 2000; Collins *et al.*, 2000). The small number of impact craters suggests that Europa's surface has undergone some kind of resurfacing by either tectonic or/and cryovolcanic processes. It is younger than the surfaces of the two other icy Galilean satellites Ganymede and Callisto. However, it is difficult to obtain absolute ages of surfaces from the number of impact craters, because the cratering rate in the outer solar system is not well known. Models suggest ages between 10 My and 1 billion years for the surface of Europa. Another possibility for assessing the age is the erosion rates by sputtering. However, the estimates of erosion rates by sputtering with which are constrained by the measurements of the H_2O escape rate during the Galileo mission. If the sputtering has been operating for more than 1 billion years, it should have erased most of the

topography; however, this is not the case for Europa. The remaining topography of Europa is consistent with an age of its surface in the range of some million years (Pappalardo *et al.*, 1999).

The disrupted blocks, also called 'rafts', have been interpreted by different models. In the thin-shelled model, the outer solid layer is less than a few kilometers thick and the surface is constantly in contact with an ocean below due to tidal cracking and disruption. Small variations in the thermal gradient cause chaos and lenticulae (Carr *et al.*, 1998; Greenberg *et al.*, 1999). In another thin-shelled model (Thomson and Delaney, 2001), the thin crust is periodically broken by faults that form in response to the tidal forces resulting from Europa's eccentric orbit around Jupiter. The blocks resulting from the faulting of the ice crust could then be entrained by the streams of the liquid ocean and could move relative to each other. Because the surface temperature is far below the freezing point, the displacement of the blocks would be limited by the rapid freezing of any liquid that would be present in between the blocks.

Another reason for the crust to be disrupted is the possibility for localized sources of heat in Europa's silicate layer. It may be conceivable that tidal forces would lead to partial melting in Europa's silicate layer and volcanism at the silicate/water interface. Such a scenario is inspired by the one that has been proposed in order to explain Io's very active volcanism. In this case, the water would be very hot during a volcanic eruption and would rise to the surface, melting the ice crust and inducing strong displacements. However, the amount of heat dissipated in Europa's silicate shell and deep interior is not well constrained and if it would allow for silicate volcanism is presently unknown.

It must be noted that volcanism is not necessarily required. Convection in Europa's silicate layer would create zones of higher temperature, and percolation of water within the silicates would create conditions favorable for hydrothermalism and subsequent displacement of hot water to the surface. Such a scenario is appealing for exobiology because the conditions at the silicate/ocean interface would be very similar to those existing on Earth's sea floor.

An alternative model suggests that the blocks and domes are related to subsolidus (solid-state) convection in the lower part of the crust (Pappalardo *et al.*, 1998; Rathbun *et al.*, 1998; McKinnon, 1999; Sotin *et al.*, 2002; Nimmo and Manga, 2002). In this model, the domes are explained by the topography resulting from the rising of hot subsolidus icy plumes forming at the

interface between the liquid layer and the icy core. It strongly suggests subsolidus convection in the icy crust. The topography of the domes is not well constrained and cannot be used to constrain the characteristics of the upwelling plumes. If this is the right explanation for the domes, it implies that the thickness of the ice crust is at least 20 km. The chaotic terrain would form by collapse of the upper crust as depressions form due to partial melt in the plume head induced by tidal heating (Sotin *et al.*, 2002).

Albeit on a smaller level as compared to Io, Europa is subject to tidal heating (Cassen *et al.*, 1979, 1980; Squyres *et al.*, 1983b; Ross and Schubert, 1987; Ojakangas and Stevenson, 1989; Hussmann *et al.*, 2002; Tobie *et al.*, 2003, 2005). The heat generated by tidal flexing can be more than one order of magnitude larger than radiogenic heating and alone could explain the presence of an ocean below the icy crust. Since the amount of ice is small, high-pressure phases cannot form and the ocean would be in contact with the silicate core, which makes Europa quite appealing from an astrobiology point of view.

There are also a few multiringed structures on Europa, such as Tyre and Callanish, which are interpreted as impact features. Study of the topographic relaxation of those features suggests a low-viscosity layer at about 5–10 km depth. This layer may be either a liquid layer or a soft ice layer. This interpretation of the formation and relaxation of impact craters is a strong argument in favor of a deep liquid ocean below an icy crust. Although a thick convecting layer of ‘warm ice’ would also be consistent with crater relaxation such a structure is unlikely because a thick low-viscosity layer would lead to vigorous tidal heating, exceeding by far the heat loss due to convection. The disequilibrium between heat production and heat flow would cause melting and thinning of the ice layer which would again imply an ocean layer at the base of the ice shell (Hussmann *et al.*, 2002).

Additional information about the internal structure of Europa comes from magnetometer measurements during the Galileo mission. Europa has a strong inductive magnetic field (see also Section 10.15.5.5) consistent with an electrically conducting interior with an outer radius equal to nearly the entire radius of the moon (Khurana *et al.*, 1998; Zimmer *et al.*, 2000). However, it does not show any evidence of a permanent dipole. The induction field can be explained by a water ocean of similar conductivity to the Earth’s oceans, provided this ocean has a thickness exceeding 10 km (Kivelson *et al.*,

2000). These measurements of an induced magnetic field provide the strongest observational evidence for the present-day existence of an ocean.

With the progress made in modeling thermal convection in planetary interiors, models describing Europa’s thermal evolution have substantially improved in the last 20 years. The main important improvements are the description of convection with complex temperature and compositional dependencies and the coupling between orbital characteristics and tidal heating, originally proposed for Io (Yoder and Peale, 1981; Greenberg, 1982; Fischer and Spohn, 1990) and later applied to Europa (Hussmann and Spohn, 2004; Tobie *et al.*, 2005).

As stated before, the amount of tidal heating is much larger for Europa than for the two Galilean satellites Ganymede and Callisto or for Saturn’s moon Titan. It may be more than an order of magnitude larger than radiogenic heating. Tidal dissipation is mainly located in the convective layer of the ice shell, that is, in the region where the temperature is close to the melting point (see Figure 11). The very effective tidal heating within a thick high-temperature layer would prevent the ice layer from freezing because with increasing ice thickness the heating rate will also increase significantly, exceeding the rate of heat transfer. The equilibrium thickness in which the tidal heating rate in the ice shell plus the radiogenic heating from the rocky mantle are balanced by the convective heat flow, is of the order of 30–40 km (Hussmann *et al.*, 2002), which is less than the total thickness of 70–200 km of the total H₂O layer derived from the gravity data, therefore suggesting the presence of an ocean.

Tidal heating in a convective shell explains the presence of an ocean within Europa even without invoking the presence of antifreeze material such as ammonia. However, it does not mean that antifreezing materials are absent. The induced magnetic field indicates that the ocean contains ions, as electrical conductors, and therefore a reduction of the melting point can be expected. The implications of this for the thermal state and ice-shell thickness on Europa have not yet been investigated in detail.

The thickness of Europa’s ice shell additionally may be subject to considerable changes with time, adjusting to the corresponding heat budget. The latter varies not only due to the decay of radioactive isotopes, which would imply an increase of thickness due to the decreasing heat source, but also due to the orbital state and the resulting tidal heating. The tidal deformation and also the tidal dissipation rate will

depend on the orbital eccentricity and mean motion (eqn [10]). Since Europa is locked in resonance with Io and Ganymede its eccentricity is forced and oscillations in both, the orbital state and the heat-production rate, may occur as a response to Io's thermal–orbital state, which is mainly controlled by tidal interaction between Io and Jupiter (Yoder, 1979; Yoder and Peale, 1981). Oscillatory phases have been proposed for Io by Ojakangas and Stevenson (1986) and by Fischer and Spohn (1990). As a consequence of such a thermal–orbital coupling the thickness of Europa's ice shell may vary between a few kilometers and a few tens of kilometers on a timescale of the order of 100 Ma (Hussmann and Spohn, 2004). The observed diversity of geological surface features on Europa may be partly explained by these time-dependent variations of ice thickness.

Because of the resonance locking, Europa's eccentricity is maintained on geological timescales. Therefore tidal heating would not cease, which is essential for any biological evolution that might have started at some point in Europa's history. Depending mainly on the tidal interaction between Jupiter and Io, tidal heating may operate on Europa for gigayears preventing the ocean from freezing and adding heat to Europa's interior. Europa is therefore a very attractive body for exobiology (Reynolds *et al.*, 1983; Greenberg *et al.*, 2000), even more because the water layer is likely to be in contact with the silicate core, allowing for complex chemical processes (Kargel *et al.*, 2000). At this interface, conditions might be similar to those at terrestrial seafloors. It has been envisaged that the conditions may be analogous to those existing at terrestrial black smokers where life develops without sunlight and where specific bacteria form. However, there is no evidence, yet, that Europa's silicate core is subject to volcanism and that tidal heating in the silicate shell is capable of producing zones of high heat flux as in the case of Io.

10.15.7.3 Titan

Titan, the largest satellite of Saturn, is unique among the moons in the solar system because of its substantial atmosphere and photochemically produced, dark-orange stratospheric haze layer that precludes direct observations of the surface. Detailed information on the surface however was obtained from the landing of the Huygens probe (Lebreton *et al.*, 2005) and by Cassini radar data which will be further collected in the mission's course (Elachi *et al.*, 2005; Stofan *et al.*, 2006). Titan's dense atmosphere that is dominated by

molecular nitrogen with a substantial component of methane (Owen, 1982; Coustenis *et al.*, 1989; Flasar *et al.*, 2005) suggests that the satellite's interior is more enriched in volatile ices such as $\text{NH}_3 \cdot \text{H}_2\text{O}$ and $\text{CH}_4 \cdot n\text{H}_2\text{O}$, which have densities similar to that of water ice (Hunten *et al.*, 1984). This is likely to be due to the lower temperature of the less-massive proto-Saturnian nebula in comparison with the proto-Jovian nebula, which favors the incorporation of a larger amount of volatiles such as ammonia and methane at the distance of Titan (20.6 Saturn radii) without significantly affecting the satellite's mean density in comparison to Ganymede and Callisto.

Furthermore, the absence of rocky inner satellites in the Saturnian system may indicate lower temperatures of the nebula during the evolution of the gaseous and dusty accretion disk from which the satellite system formed by mutual collisions of icy and rocky planetesimals (Coradini *et al.*, 1995; Owen, 2000). Based on the assumption that all nitrogen was originally supplied as NH_3 , volatile components could appear in roughly cosmic abundances in the interior of Titan with ratios of $\text{NH}_3/\text{H}_2\text{O} \approx 1/6$ and $\text{NH}_3/\text{CH}_4 \approx 1$. Corresponding water-ice–ammonia mixtures then have an ammonia content of 14 wt.%, a composition more water-rich than the peritectic composition of about 32 wt.% (Lewis, 1972; Hunten *et al.*, 1984; Stevenson, 1992).

The presence and abundance of atmospheric methane which would be lost over the age of the solar system due to photolytic dissociation (Strobel, 1974) suggests that sources of methane must be present at the surface and possibly in Titan's interior (Sotin *et al.*, 2005; Tobie *et al.*, 2006). Such a connection between interior, surface, and the dense atmosphere allowing for Titan's methane cycle is unique for the icy satellites.

Titan is intermediate between the Jovian satellites Ganymede and Callisto with respect to its radius of 2575 km and mean density of 1881 kg m^{-3} . The mean densities of Ganymede, Callisto, and Titan indicate that their interiors are composed of ice and silicates in nearly equal amounts by mass (Lupo, 1982; see above). The interior of Titan is likely to be differentiated at least into a rock + iron core and an icy mantle as a consequence of substantial accretional heating accompanied by partial outgassing and atmosphere formation (e.g., Hunten *et al.*, 1984; Grasset *et al.*, 2000). Whether or not Titan's deep interior is further differentiated like Ganymede's into an iron core and a rock mantle above it (Anderson *et al.*, 1996; Sohl *et al.*, 2002; Schubert *et al.*, 2004) is more speculative.

Depending on the amount of volatiles incorporated into the icy mantle and depending on the satellite's thermal evolution, an ammonia-rich liquid-water layer may be located between the near-surface ice-I layer and solid high-pressure ice layers at greater depth (Grasset and Sotin, 1996; Grasset *et al.*, 2000; Sohl *et al.*, 2003).

Because of Titan's close resemblance to Ganymede in mass and overall density, the evolution of the two bodies should be similar assuming identical compositions and boundary conditions. However, a number of arguments have been advanced that Titan incorporated some amount of ammonia in its interior during formation (e.g., Niemann *et al.*, 2005) and might have also incorporated large amounts of methane (Tobie *et al.*, 2006). The former implies a substantial liquid layer would exist throughout Titan's history, and the latter implies significant changes in the thermal properties of the upper ice-I crust if the methane were to force its conversion to methane clathrate hydrate (Lewis, 1971).

The ice layers crystallized upon early cooling and freezing of the pristine water-rich ammonia–water ocean that may have formed as a consequence of the blanketing effect of a hot proto-atmosphere, if accretion was completed within 10^4 – 10^5 years (Kuramoto and Matsui, 1994). It has been suggested that the ice-I layer on top of the ocean was initially cooling and thickening due to thermal conduction and became unstable to convection about 200 My after accretion (Grasset and Sotin, 1996). Some volatiles will presumably be present in the form of hydrates and clathrates thereby affecting the rheological and thermal properties of the ice layers (Hunten *et al.*, 1984; Stevenson, 1992; Tobie *et al.*, 2006). The presence of volatiles tends to reduce the freezing temperature of the ocean thereby lowering the temperature of the ice-I layer and increasing its viscosity. Larger viscosities due to temperatures far below the melting temperature then would retard convection and slow cooling of the ice-I layer, so that the subsurface ocean may have survived to the present day because of the progressing enrichment of ammonia as compared to the primordial ocean (Grasset and Sotin, 1996; Deschamps and Sotin, 2001).

Furthermore, during Ganymede's evolution into the Laplace resonance with Europa and Io, it may have been strongly heated which possibly led to complete differentiation as described in the subsection above. Titan almost certainly experienced no such event though it is subject to periodically varying tidal distortions resulting from its significant orbital

eccentricity of 0.0292 in addition to its permanent rotational flattening.

Evolution models instead are constrained to avoid excessive dissipation in Titan's interior throughout its history so as to maintain its observed present-day free eccentricity (Sohl *et al.*, 1995; Tobie *et al.*, 2005). There is considerable uncertainty in the initial eccentricity, both in general and specifically because Titan might have suffered collisions early on from one or more fragments from the breakup of proto-Hyperion (Farinella *et al.*, 1997).

Models in which Titan is volatile rich (ammonia, methane) evolve with time so as to maintain a very thin ice-I crust, typically 20 km or less, up to the last half-billion years of Titan's history, because of the low thermal conductivity and high rigidity of methane clathrate hydrate in contrast to ice-I (Tobie *et al.*, 2006). This, along with the onset of convection in the rocky core halfway through Titan's history and the gradual thickening of the crust via freezing of ice-I in the last half-billion years, provides a potential setting for repeated outgassing of methane to the surface to replace that destroyed by atmospheric photo- and charged-particle chemistry (Lorenz and Lunine, 2005). However, only indirect evidence exists for ammonia and continued methane outgassing principally through the abundances of nonradiogenic and radiogenic argon (Niemann *et al.*, 2005), and so the models must be regarded with caution.

Indeed, we do not yet know the extent to which Titan is differentiated; the dense atmosphere demands a standoff distance for the Cassini spacecraft that makes gravity measurements difficult. A volatile-poor model (Stevenson, 1992) or a model in which Titan is arranged to be partly or undifferentiated could have quite different results. The quadrupole moments of its gravity field, \mathcal{J}_2 and C_{22} , will be obtained from Doppler measurements using several flybys of the Cassini spacecraft and the MoI factor as a measure for the satellite's concentration of mass toward the center will be inferred. The measurements can be used to test the hypothesis of hydrostatic equilibrium in Titan (Rappaport *et al.*, 1997).

The orbital eccentricity of Titan causes tidal variations of the quadrupole moments from which the second-degree potential Love number k_2 will be derived. k_2 is related to the satellite's interior structure and rheological properties and will reflect the gravitational signature of a putative internal ocean. In order to obtain the value of k_2 with high accuracy, four flybys of the Cassini spacecraft will be devoted to measuring Titan's gravity field near both apokrone

and perikrone (Castillo *et al.*, 2002). k_2 may be calculated from the gravity data derived from the dedicated flybys with precision radio-tracking during which thruster operations are restricted to avoid perturbations. These will be T11, T22, T33, and T38 between Feb 2006 and Dec 2007. Some limited gravity information may be gleaned from other flybys. The expectation is that Cassini will provide MoI data akin to that for Ganymede and Callisto by the end of an extended mission (2010), but hopefully sooner.

10.15.7.4 Enceladus

As revealed by the Cassini mission, Enceladus is one of the very few planetary bodies of the solar system for which present-day activity definitely was confirmed (Porco *et al.*, 2006). Considering only the terrestrial planets and the satellites (not including comets, for instance), Enceladus is by far the smallest among these objects, with the others being the Earth, Io and Triton. That such a small moon shows signs of presently ongoing activity is unexpected, because the thermal evolution based on internal heating due to long-lived radiogenic isotopes will lead to a cold, inactive present state of the satellite, similar to the one obtained by its inner neighbor Mimas. However, the huge amount of energy required to drive the cryovolcanic plumes, which are the dominant source of Saturn's E-ring (Spahn *et al.*, 2006), and to cause the high temperatures and heat flow detected near Enceladus' South Pole (Porco *et al.*, 2006; Spencer *et al.*, 2006), still longs for an explanation. The geologically young terrain and the recently strongly modified surface in the South Pole region further suggest that Enceladus contains internal heat sources. The plumes and the maximum temperatures correlate with the so-called 'tiger stripes', parallel lineaments at the South Pole (Porco *et al.*, 2006; Spencer *et al.*, 2006). The nature of the heat-generating processes still remains unclear. One suggestion, already discussed in earlier works and motivated by surface images from the Voyager flybys, is tidal heating (Squyres *et al.*, 1983a; Poirier *et al.*, 1983; Ross and Schubert, 1989; Peale, 2003). Due to its proximity to Saturn and its elliptic orbit, Enceladus is subject to periodic tidal deformation which will lead to internal friction and thus heating. Enceladus' eccentricity is forced by Dione, which is locked in a 2:1 mean motion resonance with Enceladus, thus keeping the eccentricity finite, presently at 0.0045. Depending on the internal temperature and rheological state of the satellite, tidal friction may be sufficient to power the

observed activity. From Cassini's CIRS instrument (Composite Infrared Spectrometer), an output of 5.8 ± 1.9 GW at the south-polar hot spot was derived (Spencer *et al.*, 2006), which would be within the range of heat calculated from global tidal heating models. However, this will require low viscosities, and hence, high temperatures within the ice shell of Enceladus which are not expected on the basis of radiogenic heating, only. A liquid layer decoupling the ice shell from the rock core would additionally enhance the tidal heating rate by an order of magnitude. Global tidal heating models, however, cannot explain the asymmetry between the North Pole and South Pole of Enceladus by itself. Pockets of liquid water and local tidal heating restricted to the south-polar region would be another option. In all these models an additional process triggering vigorous tidal heating is required. A runaway process, similar to the one suggested for the rocky mantle of Io by Peale *et al.* (1979) is problematic because it would be expected rather for the innermost satellite Mimas, than for the second one, Enceladus. Under similar rheological conditions tidal heating rates are at least more than 10 times greater on Mimas, due to its proximity to Saturn and its greater orbital eccentricity (Squyres *et al.*, 1983a). The striking differences between the two satellites, with Mimas being heavily cratered and geologically inactive for billions of years and Enceladus showing an enormous internal activity, is thus even more puzzling. It is still under investigation if the higher rock content of about 50%, which was derived from a new determination of the mean density by Cassini (Porco *et al.*, 2006) compared to 20% in Mimas and the resulting higher radiogenic heating rates in Enceladus, can lead to the obviously different evolutionary paths. Possible modifications of the tidal scenario include a nonsmooth evolution into the 2:1 Enceladus–Dione mean motion resonance involving large eccentricities; a past resonance locking with the small satellite Janus enhancing the forced eccentricity of Enceladus (Lissauer *et al.*, 1984); or possible spin-orbit secondary resonances before Enceladus reached its synchronous rotational state (Wisdom, 2004). The first alternative is problematic because the evolution into resonance could also be achieved smoothly, without involving chaotic behavior and large eccentricities (Peale, 2003). However, a detailed dynamical evolution of Enceladus including tidal heating and thermal–orbital coupling has not yet been undertaken. The second alternative can only account for a heating rate five times the present one, which still would be less than

what we would get for Mimas. The third alternative would enhance the tidal heating rate by two or even three orders of magnitude if Enceladus was locked in a 3:1 secondary spin-orbit resonance (Wisdom, 2004) and could thus easily account for the difference between Enceladus and Mimas for which such a coupling cannot be expected. The problem here is that Enceladus would have to spiral inwards, contradictory to the usually assumed scenario in which a satellite moves outwards due to torques exerted by the primary. Additionally, the capture into resonance with Dione would be problematic and has not yet been investigated in detail in such a scenario. An inward movement would already indicate strong dissipation within the satellite dominating over Saturn's torques. If this happened already early in the evolution of Enceladus is unclear. All the models involving global tidal heating cannot explain the strong asymmetry between the two poles of Enceladus and to concentrate the heat sources within one relatively small region still is an open issue.

As mentioned above a liquid layer would enhance tidal heating significantly. However, radiogenic heating due to long-lived isotopes in a rock-mass fraction of 50% will be insufficient to maintain a liquid-water shell up to the present time. Although it is very likely, with regard to the energetic processes in its interior, that Enceladus is differentiated into a rock core and an H₂O layer, there is no such evidence from independent measurements. Gravity flybys to determine J_2 and C_{22} have not been performed, yet, and also the shape data is not definite. The differences of the satellite's main axes can be used to verify if it is in a hydrostatically equilibrated shape of a rotating tri-axial ellipsoid. In case of Enceladus there is some ambiguity inherent in the data, and it cannot easily be distinguished between a homogeneous structure and, for instance, a 'frozen in' differentiated structure corresponding to a higher rotation rate in the past, when Enceladus was rotating synchronously closer to Saturn (Porco *et al.*, 2006). A homogeneous structure would rule out a global ocean on Enceladus, which certainly would have implied complete differentiation. Close flybys are required to further constrain Enceladus' internal structure. It should be noted that local processes, for example, a reorientation of an upwelling plume toward the South Pole (Nimmo and Pappalardo, 2006) may contribute to the J_2 value. Thus, further flybys, equatorial and over the North Pole, would be required to reliably infer the structure of Enceladus.

The plumes observed by Cassini, which mainly consist of pure H₂O particles may be interpreted as evidence for liquid water located close to the surface (Porco *et al.*, 2006). The required heat source to melt the ice, however, still remains enigmatic (Matson *et al.*, 2007).

10.15.7.5 Rhea and Iapetus

With a radius of 764 km, Rhea is the second largest satellite in the Saturn system. The density has been determined to be 1233 kg m⁻³ (Thomas *et al.*, 2006), implying only a small amount (about 30%) of rock and a large fraction of water ice. Rhea's surface is heavily cratered. Prior to the Cassini flyby in Nov 2005, it was unclear if Rhea is differentiated and hence various possibilities had to be taken into account (Castillo-Rogez, 2006). Recent analysis of the Doppler data, however, suggests an almost homogeneous structure (Anderson and Schubert, 2007) or a weakly differentiated satellite interior (Iess *et al.*, 2006). The C_{22}/J_2 ratio is consistent with hydrostatic equilibrium within the given error bounds (Iess *et al.*, 2006). A small degree of differentiation may be due to a concentration of rocks toward the center. The possibility of Rhea being fully differentiated can be ruled out on the basis of the obtained data. This would most likely rule out the existence of a deep subsurface ocean which would have been conceivable if the satellite were differentiated and would contain at least small amounts of ammonia (Hussmann *et al.*, 2006). In the models based on the Cassini Doppler data the ice-I/ice-II phase transition occurs in the central region of the satellite. The formation of an ice-II core would be associated with a global contraction of the satellite which would have left its marks at the satellite's surface. However, showing mostly craters, the surface of Rhea is not indicative of such a process. If the formation of an ice-II core occurred early in the satellite's evolution, or if the temperature conditions inside the satellite would not have allowed for the formation of ice-II is unclear. It was also suggested that a high abundance of ammonia ice in Rhea's interior may have prevented the formation of the ice-II phase in the central regions (Prentice, 2006). Detailed imaging during the Cassini mission in combination with analysis of the gravity field may provide further insight in the internal structure of Rhea.

With respect to size and density Iapetus is comparable to Rhea. However, both satellites differ significantly in their surface characteristics as a

consequence of their individual thermal and geological evolution. Iapetus has a radius of 734.5 km and, based on its mean density of only 1088 kg m^{-3} , its rock content is even smaller than that of Rhea. Iapetus' interior is thus dominated by ice and the formation of ice-II in its central region is likely (Schubert *et al.*, 1986). Long-lived radioactive isotopes in its small rock component are insufficient to completely differentiate the satellite (Ellsworth and Schubert, 1983; Multhaup and Spohn, 2007). However, an equatorial ridge that stretches around almost the entire satellite with elevations of more than 12–20 km (Giese *et al.*, 2005; Porco *et al.* 2005) suggests that Iapetus was internally active. The ridge is heavily cratered implying internal activity in the early phases of the satellite's evolution. If there is any connection between the equatorial ridge and Cassini Regio, another unique feature on Iapetus is still under discussion (Porco *et al.*, 2005). Cassini Regio is a region with extremely low albedo centered on the leading side of the moon. This feature that causes the strong brightness variations of the satellite that is known for long from astronomical observations.

Iapetus' shape is inconsistent with its present synchronous rotation rate of almost 80 days and is best represented by an oblate spheroid (Thomas *et al.*, 2006). The difference between the equatorial and the polar radius corresponds to an equilibrium shape with a 16-h rotation rate. This suggests that Iapetus may have frozen this shape as its rotation rate was decreasing to reach the stable state of synchronous rotation (Castillo-Rogez *et al.*, 2007). If there is any connection between such a scenario and the formation of the equatorial ridge is unclear. The formation and preservation of this ridge are still enigmatic. The observations suggest that the interior was warm enough to allow for despinning and that the lithosphere was thick enough to support the mountain chain on a geologic timescale.

10.15.7.6 Triton

From the Voyager 2 flyby in 1989 Triton's radius, mass, and density were determined to be 1353.4 km (Thomas, 2000), $214 \times 10^{20} \text{ kg}$, and 2054 kg m^{-3} , respectively (Tyler *et al.*, 1989). However, on the basis of gravity measurements and from the shape solution derived from the analysis of limb profiles, the MoI cannot be inferred (Tyler *et al.*, 1989; Thomas, 2000); thus, constraints on internal structure are sparse. The high density suggests a large rock-mass fraction of about 72%, implying substantial

radiogenic heat sources in the deep interior. After Io Europa with a rock-mass fraction of 100% and 90%, respectively, Triton has the highest density of all outer planet satellites.

Because of its retrograde orbit around Neptune, it was suggested by McCord (1966) and later by McKinnon (1984) and McKinnon and Leith (1995) that Triton was captured by Neptune. In a recently proposed scenario Triton was captured as a binary object, with the companion being lost during the capture process (Agnor and Hamilton, 2006). Such models are motivated by Triton's high inclination of about 157° and its retrograde orbit, which are unique features among the large satellites in the solar system. With an eccentricity of 0.0004, Triton's orbit is almost an exact circle. Whatever the capture mechanism was like, it evolves high eccentricities shortly after capture (Ćuk and Gladman, 2005). Tidal dissipation in the satellite or gas-drag within the Neptunian subnebula (McKinnon and Leith, 1995) may have efficiently damped the satellite's eccentricity, suggesting that at least Triton's early history was characterized by intense heating. However, the young surface age of about 0.1–0.35 Ga estimated from crater counts (Stern and McKinnon, 2000; Zahnle *et al.*, 2003) suggests that Triton was geologically active, even in its recent past. This is consistent with the diversity of geological surface units imaged by Voyager 2 (Smith *et al.*, 1989). The detection of geyser-like plumes at the surface (Smith *et al.*, 1989; Soderblom *et al.*, 1990) further revealed that Triton is one of only three satellites in the solar system showing present-day eruptive activity; the other two being Io and Enceladus. However, in case of Triton the observed activity is believed to be driven by solar insolation during seasons of permanent sunlight. However, geothermal heat may additionally contribute to the required energy budget (Brown *et al.*, 1990).

The process of being captured by Neptune would lead to strong internal heating as the orbit evolved most probably implying differentiation of the satellite. This would result in a structure consisting of a rock core and an H_2O layer. The physical state of the latter remains unclear due to the not well-constrained abundance of volatiles and temperature profile in the shell. A subsurface ocean would be conceivable if small amounts of ammonia are present in the system (McKinnon and Leith, 1995). Based on thermal equilibrium models, and considering a pure ice-I layer above a water–ammonia ocean, the ice layer would be about 140–200 km thick covering an

ammonia–water ocean of about 150 km thickness (Hussmann *et al.*, 2006). Using estimates of the heat flow, required for explaining tectonic features on Triton, Ruiz (2003) estimated an ice-I layer thickness of only 20 km. Due to the lack of data on gravity and composition it is also unclear if the ocean, if it exists, would be enclosed between ice-I and high-pressure ice. Thus a ‘Europa-like’ structure consisting of a large silicate core, a water–ammonia ocean, and an ice-I shell cannot be ruled out. Such a structure would be intriguing because of its possible astrobiological consequences. Whatever the past thermal history, it has permitted outgassing of enough nitrogen and methane to produce surface ice deposits that are moved seasonally from one pole to the other and maintain a thin microbar-atmosphere of mostly nitrogen (Lunine and Nolan, 1992).

10.15.8 Future Prospects for Determining Satellite Internal Structure

At the time of writing of this chapter, the Cassini Orbiter was a bit more than halfway through its prime four-year mission around Saturn, with a number of flybys of the icy moons yet to come, and three of the four dedicated gravity passes of Titan yet to be completed. Depending on the quality of the data and success in removing the drag effects on the spacecraft of Titan’s atmosphere, determination of tidally induced shape changes on Titan from Saturn apoapse to periapse would provide a constraint on the existence of a liquid ocean in Titan’s interior. Should the Orbiter remain healthy, a two-year extended mission currently in the planning stages would be implemented, allowing for gravity mapping of Enceladus and possibly other icy moons, though this is yet to be determined.

No missions are planned to remap the shapes and internal structure of the Galilean moons, nor are there any plans to map the gravity fields of the moons of Uranus and Neptune. Juno, a polar-orbiting spacecraft to be launched to Jupiter in 2011, will provide detailed information on the internal structure and dissipation factor Q of Jupiter, which will help in understanding tidal interactions between the large moons, mainly Io, and Jupiter. New Horizons will fly by Pluto and its moon Charon in the middle of the next decade, providing limited information on internal structure. European and US scientists are interested in a Europa orbiter mission, which would

map in detail Europa’s internal structure and thickness of its ice crust (Wahr *et al.*, 2006), but this will not arrive at Jupiter any earlier than 2020. Such a mission would include flybys of the other icy moons of Jupiter, Ganymede, and Callisto. Further progress on the internal structure of icy moons can be made in laboratory studies of the relevant materials, especially with respect to the physical properties of water–ice mixed with other ices (e.g., ammonia) and improvement in theoretical modeling techniques.

References

- Agnor CB and Hamilton DP (2006) Neptune’s capture of its moon Triton in a binary-planet gravitational encounter. *Nature* 441: 192–194.
- Anderson JD, Jacobson RA, Lau EL, Moore WB, and Schubert G (2001a) Io’s gravity field, and interior structure. *Journal of Geophysical Research* 106: 32963–32970.
- Anderson JD, Jacobson RA, McElrath TP, Moore WB, Schubert G, and Thomas PC (2001b) Shape, mean radius, gravity field, and interior structure of Callisto. *Icarus* 153: 157–161.
- Anderson JD, Lau EL, Sjogren WL, Schubert G, and Moore WB (1996) Gravitational constraints on the internal structure of Ganymede. *Nature* 384: 541–543.
- Anderson JD and Schubert G (2007) Saturns satellite Rhea is a homogeneous mix of rock and ice. *Geophysical Research Letters* 34: L02202 (doi:10.1029/2006GL028100).
- Anderson JD, Schubert G, Jacobson RA, Lau EL, Moore WB, and Sjogren WL (1998) Europa’s differentiated structure, inferences from four Galileo encounters. *Science* 281: 2019–2022.
- Barr AC, Pappalardo RT, and Zhong S (2004) Convective instability in ice-I with non-Newtonian rheology: Application to the ice Galilean satellites. *Journal of Geophysical Research* 109: doi:10.1029/2004JE002296.E12008.
- Bockelée-Morvan D, Crovisier J, Mumma MJ, and Weaver HA (2004) *The composition of cometary volatiles*, In: Festou MC, Keller HU, and Weaver HA (eds.) *Comets II*, pp. 391–423. Tucson, AZ: University of Arizona Press.
- Brown RH and Clark RN (1984) Surface of Miranda: Identification of water ice. *Icarus* 58: 288–292.
- Brown RH, Clark RN, Buratti BJ, *et al.* (2006) Composition and physical properties of Enceladus’ surface. *Science* 311: 1425–1428.
- Brown RH, Johnson TV, Kirk RL, and Soderblom LA (1990) Energy sources for Triton’s geyser-like plumes. *Science* 250: 431–435.
- Calvin WM and Clark RN (1991) Modeling the reflectance spectrum of Callisto 0.25 to 4.1 μm . *Icarus* 89: 305–317.
- Canup RM and Ward WR (2002) Formation of the Galilean Satellites: Conditions of accretion. *Astronomical Journal* 124: 3404–3423.
- Canup RM and Ward WR (2006) A common mass scaling for satellite systems of gaseous planets. *Nature* 441: 834–839.
- Carr MH, Belton MJS, Chapman CR, *et al.* (1998) Evidence for a subsurface ocean on Europa. *Nature* 391: 363–365.
- Cassen P, Peale SJ, and Reynolds RT (1980) Tidal dissipation in Europa – A correction. *Geophysical Research Letters* 7: 987–988.
- Cassen P, Reynolds RT, and Peale SJ (1979) Is there liquid water on Europa? *Geophysical Research Letters* 6: 731–734.

- Castillo J, Rappaport N, Mocquet A and Sotin C (2002) Clues on Titan's internal structure from Cassini-Huygens mission. *Lunar and Planetary Science* 33 (CD-ROM), abstract 1989.
- Castillo-Rogez J (2006) Internal structure of Rhea. *Journal of Geophysical Research* 111: E11005 (doi:10.1029/2004JE002379).
- Castillo-Rogez J, Matson D, Sotin C, Johnson TV, Lunine JI, and Thomas PC (2007) Lapetus' geophysics: Rotation rate, shape and equatorial ridge. *Icarus* doi:10.1016/j.icarus.2007.02.018.
- Chizhov VE (1993) Thermodynamic properties and thermal equation of state of high-pressure ice phases. *Prikladnoi Mekhaniki Tekhnicheskoi Fiziki* (English Translation) 2: 113–123.
- Clark RN (1980) Ganymede, Europa, Callisto and Saturn's rings: Compositional analysis from reflectance spectroscopy. *Icarus* 44: 388–409.
- Clark RN, Brown RH, Jaumann R, et al. (2005) The surface composition of Phoebe as seen by the Cassini Visual and infrared mapping spectrometer. *Nature* 435: 66–69.
- Collins GC, Head JW, Pappalardo RT, and Spaun NA (2000) Evaluation of models for the formation of chaotic terrain on Europa. *Journal of Geophysical Research* 105: 1709–1716.
- Consolmagno GJ and Lewis JS (1976) Structural and thermal models of icy Galilean satellites. In: Gehrels TA (ed.) *Jupiter*, pp. 1035–1051. Tucson, AZ: University of Arizona Press.
- Consolmagno GJ and Lewis JS (1978) The evolution of icy satellite interiors and surfaces. *Icarus* 34: 280–293.
- Coradini A, Federico C, Forni O, and Magni G (1995) Origin and thermal evolution of icy satellites. *Surveys in Geophysics* 16: 533–591.
- Coustenis A, Bézard B, and Gautier D (1989) Titan's atmosphere from Voyager infrared observations. Part I: The gas composition of Titan's equatorial region. *Icarus* 80: 54–76.
- Crary FJ and Bagenal F (1998) Remanent ferromagnetism and the interior structure of Ganymede. *Journal of Geophysical Research Planets* 103: 25757–25773.
- Cruikshank DP and Brown RH (1981) The Uranian satellites: Water ice on Ariel and Umbriel. *Icarus* 45: 607–611.
- Cruikshank DP, Owen TC, Ore CD, et al. (2005) A spectroscopic study of the surfaces of Saturn's satellites: H₂O ice, tholins, and minor constituents. *Icarus* 175: 268–283.
- Cuk M and Gladman BJ (2005) Constraints on the orbital evolution of Triton. *Astronomical Journal* 626: L113–L116.
- Cyr KE, Sharp CM, and Lunine JI (1999) Effects of the redistribution of water in the solar nebula on nebular chemistry. *Journal of Geophysical Research* 104(E8): 19003–19014.
- Dartois E and d'Hendecourt L (2001) Search for ammonia ice in cold dust envelopes around YSO's. *Astronomy and Astrophysics* 365: 144–165.
- Davaille A and Jaupart C (1993) Transient high-Rayleigh-number thermal convection with large viscosity variations. *Journal of Fluid Mechanics* 253: 141–166.
- Dermott SF (1979) Shapes and gravitational moments of satellites and asteroids. *Icarus* 37: 575–586.
- Deschamps F and Sotin C (2001) Thermal convection in the outer shell of large icy satellites. *Journal of Geophysical Research* 106: 5107–5121.
- Dumoulin C, Doin M-P, and Fleitout L (1999) Heat transport in stagnant lid convection with temperature- and pressure-dependent Newtonian and non-Newtonian rheology. *Journal of Geophysical Research* 104: 12759–12777.
- Durham WB and Stern LA (2001) Rheological properties of water ice – Applications to satellites of the outer planets. *Annual Review of Earth and Planetary Sciences* 29: 295–330.
- Elachi C, Wall S, Allison M, et al. (2005) First views of the surface of Titan from the Cassini RADAR. *Science* 308: 970–974.
- Ellsworth K and Schubert G (1983) Saturn's icy satellites – Thermal and structural models. *Icarus* 54: 490–510.
- Estrada PR and Mosqueira I (2006) A gas-poor planetesimal capture model for the formation of giant planet satellite systems. *Icarus* 181: 486–509.
- Farinella P, Mazari F, and Matteoli S (1997) The disruption of Hyperion and the origin of Titan's atmosphere. *Astronomical Journal* 113: 2312–2316.
- Fischer H-J and Spohn T (1990) Thermal-orbital histories of viscoelastic models of Io (J1). *Icarus* 83: 39–65.
- Flasar F, Achterberg RK, Conrath BJ, et al. (2005) Titan's atmospheric temperatures, winds and composition. *Science* 308: 975–978.
- Giese B, Denk T, Neukum G, Porco CC, Roatsch T, and Wagner R (2005) The topography of Iapetus' leading side. *Bulletin of the American Astronomical Society* 37: 728.
- Goldsby DL and Kohlstedt DL (2001) Superplastic deformation of ice: Experimental observations. *Journal of Geophysical Research* 106: 11017–11030.
- Grasset O and Pargamin J (2005) The ammonia water system at high pressures: Implications for the methane of Titan. *Planetary and Space Science* 53: 371–384.
- Grasset O and Parmentier EM (1998) Thermal convection in a volumetrically heated, infinite Prandtl number fluid with strongly temperature-dependent viscosity: Implications for planetary evolution. *Journal of Geophysical Research* 103: 18171–18181.
- Grasset O and Sotin C (1996) The cooling rate of a liquid shell in Titan's interior. *Icarus* 123: 101–112.
- Grasset O, Sotin C, and Deschamps F (2000) On the internal structure and dynamics of Titan. *Planetary and Space Science* 48: 617–636.
- Greeley R, Sullivan R, Klemaszewski J, et al. (1998) Europa: Initial Galileo geological observations. *Icarus* 135: 4–24.
- Greenberg R (1982) Orbital evolution of the Galilean satellites. In: Morrison D (ed.) *Satellites of Jupiter*, pp. 65–92. Tucson, AZ: University of Arizona Press.
- Greenberg R, Croft SK, Janes DM, et al. (1991) Miranda. In: Bergstrahl JT, Miner ED, and Matthews MS (eds.) *Uranus*, pp. 693–735. Tucson: University of Arizona Press.
- Greenberg R, Geissler P, Hoppa G, et al. (1998) Tectonic processes on Europa: Tidal stresses, mechanical response, and visible features. *Icarus* 135: 64–78.
- Greenberg R, Geissler P, Tufts BR, and Hoppa GV (2000) Habitability of Europa's crust: The role of tidal-tectonic processes. *Journal of Geophysical Research* 105(E7): 17551–17562.
- Greenberg R, Hoppa GV, Tufts BR, Geissler P, Riley J, and Kadel S (1999) Chaos on Europa. *Icarus* 141: 263–286.
- Grevesse N, Asplund M, and Sauval AJ (2005) The new solar chemical composition. In: Alecian G, Richard O, and Vauclair S (eds.) *EAS Publication series Vol 9, Element Stratification in Stars, 40 Years of Atomic Diffusion*, pp. 1–10. Les Ulis, France: EDP Sciences Publication.
- Hibbits CA, Pappalardo RT, Hansen GB, and McCord TB (2003) Carbon dioxide on Ganymede. *Journal of Geophysical Research* 108(E5): 5036 (doi:10.1029/2002JE001956).
- Hobbs PV (1974) *Ice Physics*. Oxford: Clarendon Press.
- Hogenboom DL, Kargel JS, Consolmagno GJ, Holden TC, Lee L, and Buyounouski M (1997) The ammonia–water system and the chemical differentiation of icy satellites. *Icarus* 128: 171–180.
- Hoppa GV, Tufts BR, Greenberg R, and Geissler PE (1999) Formation of cycloidal features on Europa. *Science* 285: 1899–1902.

- Hubbard WB and Anderson JD (1978) Possible flyby measurements of Galilean satellite interior structure. *Icarus* 33: 336–341.
- Hunten DM, Tomasko MG, Flasar FM, Samuelson RE, Strobel DF, and Stevenson DJ (1984) Titan. In: Gehrels T and Matthews MS (eds.) *Saturn*, pp. 671–759. Tucson, AZ: University of Arizona Press.
- Hussmann H, Sohl F, and Spohn T (2006) Subsurface oceans of medium-sized outer planet satellites and large trans-Neptunian objects. *Icarus* 185: 258–273.
- Hussmann H and Spohn T (2004) Thermal-orbital evolution of Io and Europa. *Icarus* 171: 391–410.
- Hussmann H, Spohn T, and Wieczorkowski K (2002) Thermal equilibrium states of Europa's ice shell: Implications for internal ocean thickness and surface heat flow. *Icarus* 156: 143–151.
- Less L, Rappaport NJ, Tortora P, et al. (2006) Gravity field and interior of Rhea from Cassini data analysis. *Icarus* (in press).
- Jacobson RA (2004) The orbits of the major Saturnian satellites and the gravity field of Saturn from spacecraft and Earth-based observations. *Astronomical Journal* 128: 492–501.
- Jacobson RA, Antreasian PG, Bordi JJ, et al. (2006) The gravity field of the saturnian system from satellite observations and spacecraft tracking data. *Astronomical Journal* 132: 2520–2526.
- Johnson TV and Lunine JI (2005) Saturn's moon Phoebe as a captured body from the outer solar system. *Nature* 435: 69–71.
- Kargel JS, Kaye JZ, Head JW, et al. (2000) Europa's crust and ocean: Origin, composition, and the prospects for life. *Icarus* 148: 226–265.
- Kaula WM (1968) *Space Science Text Series: An Introduction to Planetary Physics: The Terrestrial Planets*. New York: Wiley.
- Khurana KK, Kivelson MG, Stevenson DJ, et al. (1998) Induced magnetic fields as evidence for subsurface oceans in Europa and Callisto. *Nature* 395: 777–780.
- Kirk RL and Stevenson DJ (1987) Thermal evolution of a differentiated Ganymede and implications for surface features. *Icarus* 69: 91–134.
- Kivelson MG, Khurana KK, Krishan K, et al. (2000) Galileo magnetometer measurements: A stronger case for a subsurface ocean at Europa. *Science* 289: 1340–1343.
- Kivelson MG, Khurana KK, and Volwerk M (2002) The permanent and inductive magnetic moments of Ganymede. *Icarus* 157: 507–522.
- Kivelson MG, Warnecke J, Bennett L, et al. (1998) Ganymede's magnetosphere: Magnetometer overview. *Journal of Geophysical Research Planets* 103: 19963–19972.
- Klinger J (1981) Some consequences of a phase transition of water ice on the heat balance of comet nuclei. *Icarus* 47: 320–324.
- Kuramoto K and Matsui T (1994) Formation of a hot proto-atmosphere on the accreting giant icy satellite: Implications for the origin and evolution of Titan, Ganymede, and Callisto. *Journal of Geophysical Research* 99: 21183–21200.
- Kuskov OL and Kronrod VA (2005) Internal structure of Europa and Callisto. *Icarus* 177: 550–569.
- Lebreton J-P, Witasse O, Sollazzo C, et al. (2005) An overview of the descent and landing of the Huygens probe on Titan. *Nature* 438: 758–764.
- Leliwa-Kopystyński J, Maruyama M, and Nakajima T (2002) The water-ammonia phase diagram up to 300MPa: Application to icy satellites. *Icarus* 159: 518–528.
- Lewis JS (1971) Satellites of the outer planets: Their physical and chemical nature. *Icarus* 15: 175–185.
- Lewis JS (1972) Low temperature condensation from the solar nebula. *Icarus* 16: 241–252.
- Lindal GF, Wood GE, Hotz HB, Sweetnam DN, Eshleman VR, and Tyler GL (1983) The atmosphere of Titan – An analysis of the Voyager 1 radio occultation measurements. *Icarus* 53: 348–363.
- Lissauer JJ, Peale SJ, and Cuzzi JN (1984) Ring torque on Janus and the melting of Enceladus. *Icarus* 58: 159–168.
- Lorenz R and Lunine J (2005) Titan's surface before Cassini. *Planetary and Space Science* 53: 557–576.
- Lunine JI and Nolan MC (1992) A massive early atmosphere on Triton. *Icarus* 100: 221–234.
- Lunine JI and Stevenson DJ (1982) Formation of the Galilean satellites in a gaseous nebula. *Icarus* 52: 14–39.
- Lunine JI and Stevenson DJ (1987) Clathrate and ammonia hydrates at high pressure – Application to the origin of methane on Titan. *Icarus* 70: 61–77.
- Lupo MJ (1982) Mass-radius relationships in icy satellites after Voyager. *Icarus* 52: 40–53.
- Matson DL, Castillo JC, Lunine JI, and Johnson TV (2007) Enceladus' plume: Compositional evidence for a hot interior. *Icarus* 187: 569–573.
- McCord TB (1966) Dynamical evolution of the Neptunian system. *Astronomical Journal* 71: 585.
- McKinnon WB (1984) On the origin of Triton and Pluto. *Nature* 311: 355–358.
- McKinnon WB (1999) Convective instability in Europa's floating ice shell. *Geophysical Research Letters* 26: 951–954.
- McKinnon WB (2006) On convection in ice I shells of outer solar system bodies, with detailed application to Callisto. *Icarus* 183: 435–450.
- McKinnon WB and Leith AC (1995) Gas drag and the orbital evolution of a captured Triton. *Icarus* 118: 392–413.
- Merk R, Breuer D, and Spohn T (2002) Numerical modelling of ^{26}Al -induced radioactive melting of asteroids considering accretion. *Icarus* 159: 183–191.
- Mitri G and Showman AP (2005) Convective-conductive transitions and sensitivity of a convecting ice shell to perturbations in heat flux and tidal heating rate: Implications for Europa. *Icarus* 177: 447–460.
- Moore WB and Schubert G (2000) The tidal response of Europa. *Icarus* 147: 317–319.
- Moore WB and Schubert G (2003) The tidal response of Ganymede and Callisto with and without liquid water oceans. *Icarus* 166: 223–226.
- Mousis O, Gautier D, and Bockelée-Morvan D (2002) An evolutionary turbulent model of Saturn's subnebula, implications on the origin of the atmosphere of Titan. *Icarus* 156: 162–175.
- Mueller S and McKinnon WB (1988) Three-layered models of Ganymede and Callisto: Compositions, structures and aspects of evolution. *Icarus* 76: 437–464.
- Multhaup K and Spohn T (2007) Stagnant lid convection in the mid-sized icy satellites of Saturn. *Icarus* 186: 420–435.
- Nagel K, Breuer D, and Spohn T (2004) A model for the interior structure, evolution, and differentiation of Callisto. *Icarus* 169: 402–412.
- Neubauer FM (1999) Alfvén wings and electromagnetic induction in the interiors: Europa and Callisto. *Journal of Geophysical Research* 104(A12): 28671.
- Niemann HB, Atreya SK, Bauer SJ, et al. (2005) The abundances of constituents of Titan's atmosphere from the GCMS instrument on the Huygens probe. *Nature* 438: 779–784.
- Nimmo F and Manga M (2002) Causes, characteristics and consequences of convective diapirism on Europa. *Geophysical Research Letters* 29: 24.
- Nimmo F and Pappalardo RT (2006) Diapir-induced reorientation of Saturn's moon Enceladus. *Nature* 441: 614–616.

- Ojakangas GW and Stevenson DJ (1986) Episodic volcanism of tidally heated satellites with application to Io. *Icarus* 66: 341–358.
- Ojakangas GW and Stevenson DJ (1989) Thermal state of an ice shell on Europa. *Icarus* 81: 220–241.
- Okin CB, Wasserman LH, and Franz OG (2003) The mass ratio of Charon to Pluto from Hubble space telescope astrometry with the fine guidance sensors. *Icarus* 164: 254–259.
- Owen T (1982) The composition and origin of Titan's atmosphere. *Planetary and Space Science* 30: 833–838.
- Owen T (2000) On the origin of Titan's atmosphere. *Planetary and Space Science* 48: 747–752.
- Pappalardo RT, Belton MJS, Breneman HH, et al. (1999) Does Europa have a subsurface ocean?: Evaluation of the geological evidence. *Journal of Geophysical Research* 104: 24015–24055.
- Pappalardo RT, Head J, Greeley WR, et al. (1998) Geologic evidence for solid-state convection in Europa's ice shell. *Nature* 391: 365–368.
- Peale SJ (1986) Orbital resonances, unusual configurations and exotic rotation states among planetary satellites. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 159–223. Tucson, AZ: University of Arizona Press.
- Peale SJ (2003) Tidally induced volcanism. *Celestial Mechanics and Dynamical Astronomy* 87: 129–155.
- Peale SJ, Cassen P, and Reynolds RT (1979) Melting of Io by tidal dissipation. *Science* 203: 892–894.
- Peale SJ and Lee MH (2002) A primordial origin of the Laplace relation among the Galilean satellites. *Science* 298: 593–597.
- Poirier JP, Boloh L, and Chambon P (1983) Tidal dissipation in small viscoelastic ice moons – The case of Enceladus. *Icarus* 55: 218–230.
- Porco CC, Baker E, Barbara J, et al. (2005) Cassini imaging science: Initial results on Phoebe and Iapetus. *Science* 307: 1237–1242.
- Porco CC, Helfenstein P, Thomas PC, et al. (2006) Cassini observes the active south pole of Enceladus. *Science* 311: 1393–1401.
- Prentice AJR (2006) Saturn's icy moon Rhea: A prediction for its bulk chemical composition and physical structure at the time of the Cassini spacecraft first flyby. *Publications of the Astronomical Society of Australia* 23: 1–11.
- Rappaport N, Bertotti B, Giampieri G, and Anderson JD (1997) Doppler measurements of the quadrupole moments of Titan. *Icarus* 126: 313–323.
- Rappaport NJ, Giampieri G, and Anderson JD (2001) Perturbations of a spacecraft orbit during a hyperbolic flyby. *Icarus* 150: 168–180.
- Rathbun JA, Musser GS, and Squyres SW (1998) Ice diapers on Europa: Implications for liquid water. *Geophysical Research Letters* 25: 4157–4160.
- Reynolds RT, Squyres SW, Colburn DS, and McKay CP (1983) On the habitability of Europa. *Icarus* 56: 246–254.
- Riley J, Hoppa GV, Greenberg R, Tufts BR, and Geissler P (2000) Distribution of chaotic terrain on Europa. *Journal of Geophysical Research* 105: 22599–22615.
- Ross MN and Schubert G (1987) Tidal heating in an internal ocean model of Europa. *Nature* 325: 132–134.
- Ross MN and Schubert G (1989) Viscoelastic models of tidal heating in Enceladus. *Icarus* 78: 90–101.
- Ruiz J (2001) The stability against freezing of an internal liquid-water ocean on Callisto. *Nature* 412: 409–411.
- Ruiz J (2003) Heat flow and depth to a possible internal ocean on Triton. *Icarus* 166: 436–439.
- Schubert G, Anderson JD, Spohn T, and McKinnon WB (2004) Interior composition, structure, and dynamics of the Galilean Satellites. In: Bagenal F, Dowling TE, and McKinnon WB (eds.) *Jupiter. The Planet, Satellites, and Magnetosphere*, pp. 281–306. Cambridge, UK: Cambridge University Press.
- Schubert G, Cassen P, and Young RE (1979) Subsolidus convective cooling histories of terrestrial planets. *Icarus* 38: 192–211.
- Schubert G, Spohn T, and Reynolds RT (1986) Thermal histories, compositions and internal structures of the moons of the solar system. In: Burns JA and Matthews MS (eds.) *Satellites*, pp. 224–292. Tucson, AZ: Arizona University Press.
- Schubert G, Stevenson DJ, and Ellsworth K (1981) Internal structures of the Galilean satellites. *Icarus* 47: 46–59.
- Schubert G, Zhang K, Kivelson MG, and Anderson JD (1996) The magnetic field and internal structure of Ganymede. *Nature* 384: 544–545.
- Segatz M, Spohn T, Ross MN, and Schubert G (1988) Tidal dissipation, surface heat flow and figures of viscoelastic models of Io. *Icarus* 75: 187–206.
- Showman AP and Han L (2004) Numerical simulation of convection in Europa's ice shell: Implications for surface features. *Journal of Geophysical Research* 109(E1): E01010 (doi: 10.1029/2003JE002103).
- Showman AP and Malhotra R (1997) Tidal evolution into the Laplace resonance and the resurfacing of Ganymede. *Icarus* 127: 93–111.
- Smith BA, Soderblom LA, and Banfield D (1989) Voyager 2 at Neptune—Imaging science results. *Science* 246: 1422–1449.
- Soderblom LA, Becker TL, Kieffer SW, Brown RH, Hansen CJ, and Johnson TV (1990) Triton's geyser-like plumes – Discovery and basic characterization. *Science* 250: 410–415.
- Sohl F, Hussmann H, Schwentker B, and Spohn T (2003) Interior structure and tidal Love numbers of Titan. *Journal of Geophysical Research* 108(E12): 5130 (doi: 10.1029/2003JE002044).
- Sohl F, Sears WD, and Lorenz R (1995) Tidal dissipation on Titan. *Icarus* 115: 278–294.
- Sohl F, Spohn T, Breuer D, and Nagel K (2002) Implications from Galileo observations on the interior structure and chemistry of the Galilean satellites. *Icarus* 157: 104–119.
- Solomatov VS (1995) Scaling of temperature- and stress-dependent viscosity convection. *Physics of Fluids* 7: 266–274.
- Sotin C, Grasset O, and Beauchesne S (1998) Thermodynamical properties of high pressure ices: Implications for the dynamics and internal structure of large icy satellites. In: Schmitt B, De Bergh M, and Festou C (eds.) *Solar System Ices*, pp. 79–96. Norwell, MA: Kluwer Academic.
- Sotin C, Head JW, and Tobie G (2002) Europa: Tidal heating of upwelling thermal plumes and the origin of lenticulae and chaos melting. *Geophysical Research Letters* 29(8): 74–1–74–4.
- Sotin C, Jaumann R, Buratti BJ, et al. (2005) Release of volatiles from a possible cryovolcano from near-infrared imaging of Titan. *Nature* 435: 786–789.
- Sotin C and Tobie G (2004) Internal structure and dynamics of the large icy satellites. *Comptes Rendus de l'Académie des Sciences* 335: 769–780.
- Spahn F, Schmidt J, Albers N, et al. (2006) Cassini dust measurements at Enceladus and implications for the origin of the E-ring. *Science* 311: 1416–1418.
- Spaun NA, Head JW, Collins GC, Prockter LM, and Pappalardo RT (1998) Conamara Chaos region, Europa: Reconstruction of mobile polygonal ice blocks. *Geophysical Research Letters* 25: 4277–4280.
- Spencer JR, Pearl JC, Segura M, et al. (2006) Cassini encounters Enceladus: Background and the discovery of a south polar hot spot. *Science* 311: 1401–1405.

- Spoohn T (1997) Tides of Io. In: Wilhelm H, Zürn W, and Wenzel H-G (eds.) *Tidal Phenomena*, pp. 345–377. Heidelberg: Springer.
- Spoohn T and Schubert G (2003) Oceans in the icy Galilean satellites of Jupiter? *Icarus* 161: 456–467.
- Squyres SW, Reynolds RT, and Cassen PM (1983b) Liquid water and active resurfacing on Europa. *Nature* 301: 225–226.
- Squyres SW, Reynolds RT, Cassen PM, and Peale SJ (1983a) The evolution of Enceladus. *Icarus* 53: 319–331.
- Stern SA and McKinnon WB (2000) Triton's surface age and impactor population revisited in light of Kuiper belt fluxes: Evidence for small Kuiper belt objects and recent geological activity. *Astronomical Journal* 119: 945–952.
- Stevenson DJ (1992) The interior of Titan. In: *Proceedings of Symposium on Titan*, ESA SP-338. ESA, Noordwijk, The Netherlands, pp. 29–33.
- Stevenson DJ (2003) Planetary magnetic fields. *Earth and Planetary Science Letters* 208: 1–11.
- Stevenson DJ and Lunine JI (1988) Rapid formation of Jupiter by diffuse redistribution of water vapour in the solar nebula. *Icarus* 75: 146–155.
- Stofan ER, Lunine JI, Lopes R, et al. (2006) Mapping of Titan: Results from the first Titan radar passes. *Icarus* 185: 443–456.
- Strobel DF (1974) The photochemistry in the atmosphere of Titan. *Icarus* 21: 466–470.
- Tholen DJ and Buie MW (1990) Further analysis of the Pluto–Charon mutual event observations. *Bulletin of the American Astronomical Society* 22: 1129.
- Thomas PC (2000) The shape of Triton from limb profiles. *Icarus* 148: 587–588.
- Thomas PC, Helfenstein P, Veverka J, et al. (2006) Sizes, shapes, relaxation states and interior configurations of icy saturnian satellites. *American Astronomical Society, DPS-meeting #38*, abstract 69.02.
- Thomson RE and Delaney JR (2001) Evidence for a weakly stratified European ocean sustained by seafloor heat flux. *Journal of Geophysical Research* 106: 12355–12365.
- Tobie G, Choblet G, and Sotin C (2003) Tidally heated convection: Constraints on Europa's ice shell thickness. *Journal of Geophysical Research* 108: doi:10.1029/2003JE002099.
- Tobie G, Lunine J, and Sotin C (2006) Episodic outgassing as the source of atmospheric methane on Titan. *Nature* 440: 61–64 (doi:10.1038/nature04497).
- Tobie G, Mocquet A, and Sotin C (2005) Tidal dissipation within large icy satellites: Europa and Titan. *Icarus* 177: 534–549.
- Tomaso M, Archinal B, Becker T, et al. (2005) Rain, winds, and haze during the Huygens probe descent to Titan's surface. *Nature* 438: 765–778.
- Tyler GL, Sweetnam DN, Anderson JD, et al. (1989) Voyager radio science observations of Neptune and Triton. *Science* 246: 1466–1473.
- Wahr JM, Zuber MT, Smith DE, and Lunine JI (2006) Tides on Europa and the thickness of Europa's icy shell. *Journal of Geophysical Research* 111: E12 (doi:10.1029/2006JE002729).
- Waite JH, Jr., Combi M, Ip W-H, et al. (2006) Cassini ion and neutral mass spectrometer: Enceladus plume composition and structure. *Science* 311: 1419–1422.
- Weertman J (1983) Creep deformation of ice. *Annual Review of Earth and Planetary Sciences* 11: 215–240.
- Wisdom J (2004) Spin-orbit secondary resonance dynamics of Enceladus. *Astronomical Journal* 128: 484–491.
- Yoder CF (1979) How tidal heating in Io drives the Galilean orbital resonance locks. *Nature* 279: 767–770.
- Yoder CF and Peale SJ (1981) The tides of Io. *Icarus* 47: 1–35.
- Zahnle K, Schenk P, Levison H, and Dones L (2003) Cratering rates in the outer solar system. *Icarus* 163: 263–289.
- Zharkov VN, Leontjev VV, and Kozenko AV (1985) Models, figures, and gravitational moments of the Galilean satellites of Jupiter and icy satellites of Saturn. *Icarus* 61: 92–100.
- Zimmer C, Khurana KK, and Kivelson MG (2000) Subsurface oceans on Europa and Callisto: Constraints from Galileo magnetometer observations. *Icarus* 147: 329–347.

10.16 Pluto, Charon, and the Kuiper Belt Objects

S. A. Stern, J. Wm. Parker, and C. B. Olkin, Southwest Research Institute, Boulder, CO, USA

© 2007 Elsevier B.V. All rights reserved.

10.16.1	Overviews	541
10.16.1.1	The Pluto System	541
10.16.1.2	Kuiper Belt	542
10.16.1.3	Kuiper Belt Objects	544
10.16.1.4	On KBO Nomenclature	544
10.16.2	Environment	545
10.16.2.1	Insolation	545
10.16.2.2	Collisions	545
10.16.2.3	Erosional Interactions with the Interstellar Medium	546
10.16.2.4	Radiation Processes	546
10.16.3	Physical Properties	546
10.16.3.1	Albedos	546
10.16.3.2	Rotational Properties and Light Curves	547
10.16.3.3	Colors	548
10.16.3.4	Surface Composition	550
10.16.3.5	Atmospheres	552
10.16.3.5.1	Pluto's atmosphere	553
10.16.3.5.2	Other KBOs	555
10.16.3.6	Densities	556
10.16.3.7	Interior Models	556
10.16.4	Origin	557
10.16.5	Future Observational Goals and Prospects	558
10.16.5.1	Goals	558
10.16.5.2	Future Ground-Based Observatories	559
10.16.5.3	Future Space-Based Observations	560
10.16.5.4	Space Missions to the KB and Centaurs	560
References		561

10.16.1 Overviews

It is now widely recognized that Pluto's discovery was a harbinger of the discovery of the Kuiper Belt (KB), the third major architectural zone of our solar system, after the terrestrial planet and giant planet zones. It is less widely recognized, but nonetheless the case, that Pluto has over time also been a harbinger of dwarf planets, of slow rotation periods, of bodies containing exotic surface volatiles, of bodies trapped in mean motion resonance (MMR) with Neptune, and of satellites that are now seen to be common among Kuiper Belt objects (KBOs). Some day, Pluto may be seen as a harbinger of KBO atmospheres.

In this section, we provide an overview of what is known about the Pluto system, the KB, and KBOs. In subsequent sections, more detail is provided.

10.16.1.1 The Pluto System

The discovery of Pluto by Tombaugh in 1930 opened a chapter in planetary science that is still unfolding. Almost immediately after this discovery was announced, it was recognized that this new planet was highly unusual, owing to its small size, and its highly elliptical, highly inclined orbit. As a result, struggles to place Pluto in the context of the architecture of our solar system continued for many decades. The range of conjecture concerning Pluto spanned the space from it being an ejected asteroid, to an escaped satellite, to a member of a distant debris belt related to planet formation (e.g., see Stern *et al.*, 1997).

However, owing to limitations in observing and computing technology, little progress was made in understanding Pluto's context before the 1980s, when

serious discussion of the KB (Fernandez and Ip, 1984; Duncan *et al.*, 1988; Stern, 1991) and the population content of 1000-km-sized bodies in the deep outer solar system was seriously examined from a quantitative standpoint.

Similarly, owing to its faintness and small angular size on the sky, which put its study well beyond the capabilities of mid-twentieth century technology, little progress regarding Pluto's nature was made before the 1950s. However, from 1955 to the present, the increasing pace of technological advances in astronomical telescopes and instrumentation, combined with advances in perspective drawn from the initial reconnaissance of the terrestrial and giant planets, and the discovery of the KB, led to an increasingly full picture of Pluto and its context. Major advances achieved in this era included the following:

- Discovery of Pluto's 6.4-day rotation period ([Walker and Hardie, 1955](#)).
- Discovery of Pluto's 3:2 mean motion heliocentric resonance with Neptune ([Cohen and Hubbard, 1965](#)).
- First crude spectrum, revealing a reddish surface color ([Fix *et al.*, 1970](#)).
- CH₄-ice detected ([Cruikshank *et al.*, 1976](#)).
- Charon was discovered, yielding a system mass ([Christy and Harrington, 1978](#)).
- Stellar occultation revealed that Charon's radius is close to 600 km ([Walker, 1980](#)).
- Onset of a rich series of mutual occultation events between Pluto and Charon ([Binzel *et al.*, 1985](#)).
- Determination of reliable radii for Pluto and Charon ([Tholen *et al.*, 1986](#)), in turn yielding a 2 g cm⁻³ average system density.
- Water ice was discovered on Charon ([Marcialis *et al.*, 1987](#)).
- Pluto's atmosphere was revealed by stellar occultation ([Elliot *et al.*, 1989](#)).
- Mutual events produced evidence for polar caps ([Binzel, 1989](#)).
- Arguments that ice dwarf planets like Pluto were common in the early solar system ([Stern, 1991](#)).
- Discovery of the KB ([Jewitt and Luu, 1992](#)).
- Discovery of N₂ and CO ices on Pluto ([Owen *et al.*, 1993](#)).
- Hubble Space Telescope (HST) imaging revealed Pluto's highly variegated surface ([Stern *et al.*, 1997](#)).
- Detection of CH₄ in Pluto's atmosphere ([Young *et al.*, 1997](#)).
- Discovery of Pluto's small satellites Nix and Hydra ([Weaver *et al.*, 2006; Stern *et al.*, 2006](#)).

With many of these facts in hand, by the late 1980s it was clear that Pluto was a very different kind of world from either the terrestrial planets or the giant planets, replete with multiple surface volatiles, a complex, escaping, seasonally variable atmosphere, and a highly variegated surface. Moreover, Pluto was seen to be a binary planet, that is, an object with a satellite so large, and so close, that the system barycenter lies outside the primary in free space between the planet and the moon. By the early 1990s, it had also been argued that Pluto was not a misfit among the planets, but the first of a large population of 1000-km-sized KB and more distant bodies in a population that likely include objects of sizes extending up to Earth mass or somewhat larger ([Stern, 1991](#)).

10.16.1.2 Kuiper Belt

The existence of the KB was first predicted by mid-twentieth century astronomers such as Kenneth Edgeworth (1943, 1949) and Gerard Kuiper (1951). These and other astronomers of the 1930s, 1940s, and 1950s postulated that a debris belt of material left over from planetary formation might orbit the Sun beyond Neptune. However, the faint object detection technology of the mid-twentieth century was too primitive to give astronomers real hope of finding bodies out there – they were simply too faint to be found.

By the late 1980s, however, dynamicists found strong evidence in the inclination distribution of the Jupiter family comets that they are coming from a disk-like reservoir just beyond Neptune's orbit ([Duncan *et al.*, 1988](#)). This dynamical evidence was bolstered by the discovery of 2060 Chiron, orbiting in a short-lived, dynamically chaotic orbit, between Saturn and Uranus ([Kowal, 1977](#)). Unfortunately, although Kowal's survey data set did detect several large KBOs, his manual detection technique failed to identify them as trans-Neptunian bodies, so they remained undiscovered for another 20+ years.

As a result of the discovery of Chiron and computer modeling that argued strongly for a disk-like reservoir of comets beyond the giant planets (e.g., [Duncan *et al.*, 1987](#)), a number of searches were begun in the late 1980s for the belt of material that Kuiper predicted. The first KBO was subsequently discovered ([Jewitt and Luu, 1992](#)). This object, designated 1992QB₁, is more than 1000 times fainter than Pluto, and probably some 10–20 times smaller in radius.

Over 1200 other KBOs and centaurs had been discovered by late 2006, with estimated diameters ranging from 30 to perhaps 2700 km (e.g., Bernstein *et al.*, 2004; Bertoldi *et al.*, 2006). Based on the large fraction of sky left to be searched and the sky density of faint, distant objects being found in deep charge-coupled device (CCD) images, it is estimated that over 100 000 KBOs with diameters more than 50 km may orbit in the KB. From the orbital distribution of KBOs, the KB has been revealed to be a disk- or belt-like structure that stretches from 30 to at least 55 astronomical units (AU) from the Sun. It is now recognized that the KB is a far greater collection of objects, particularly large objects, than the asteroid belt between Mars and Jupiter.

Collisional processes are known to play a key role in the KB (e.g., Stern, 1995, 1996; Farinella and Davis, 1996; Davis and Farinella, 1997; Durda and Stern, 2000). A significant result of collisional modeling is that KBOs smaller than ~ 50 km in diameter cannot have survived the collisional bombardment of the current-day KB over time and therefore must be much younger than the age of the solar system. This implies that the Jupiter family comets, which have their source region in the KB, are recently produced chips off larger, ancient KBOs.

Based on the sizes and orbits of KBOs, it appears that the birthplace of the KBOs was well on its way to growing one or more large planets, perhaps even something the size of the Earth, or even Neptune, when the growth process was interrupted (e.g., Stern and Colwell, 1997a, 1997b; Farinella *et al.*, 2000). Perhaps the formation or transport of Neptune near 30 AU interrupted this growth.

Numerous excellent reviews of the structure and dynamics of the KB have been written (e.g., Morbidelli, 2006). Here we give only a brief recapitulation of the primary structure and orbital populations of the KB, in order to set the context for our discussions concerning KBOs themselves.

The KB consists of two broad dynamical populations, the classical KB (CKB) and the scattered KB (SKB). Both populations involve bodies on orbits that lie beyond Neptune. Former KBOs, which currently penetrate deeply into the giant planets' region owing to dynamical loss from the CKB and SKB, are referred to as centaurs. Centaur bodies have short (e.g., $10^{6.5} - 10^{7.5}$ years) orbital lifetimes that typically end with a strong scattering event that ejects them back outside the giant planet region.

The classical belt contains objects that orbit beyond Neptune but generally within ~ 60 AU.

Both Neptune-resonant and Neptune-nonresonant classical KBOs (CKBOs) are recognized. Those on resonant orbits are often referred to as plutinos, in honor of Pluto, the first object discovered on such an orbit type. Although Pluto orbits in the 3:2 Neptune MMR, plutinos have also been found in the 2:1, 5:3, and 7:4 Neptune MMRs as well.

The MMR orbits of the plutinos provide strong evidence for the migration of Neptune, probably by many astronomical units, early in the history of the solar system (e.g., Malhotra, 1993; Morbidelli *et al.*, 2005). Further, the broad distribution of CKBO orbital inclinations and the presence of the scattered KBOs (SKBOs) indicate a violent dynamical history there. This might involve a passing star, large migrating planetary embryos, or sweeping secular resonances during the time that the KBOs were forming (see, e.g., Morbidelli, 2006).

Estimates of the mass of the current-day KB fall in the range of $10^{-2} M_{\text{Earth}}$ to $10^{-1} M_{\text{Earth}}$, with crudely equal proportions in the CKB and SKB (Morbidelli, 2006). Less than 0.1% of this mass is at any time orbiting closer as Centaurs (Morbidelli, 2006).

The fact that large bodies hundreds and thousands of kilometers across orbit in the KB provides inescapable evidence that these bodies formed in an environment with approximately $10^2 - 10^3$ times higher surface mass density than is present today (e.g., Stern, 1995, 1996; Farinella *et al.*, 2000; Kenyon and Bromley, 2001). This in turn suggests that either the KB region was formerly far more massive and was subsequently dynamically and/or collisionally eroded (e.g., Stern and Colwell, 1997a, 1997b) or that the large bodies in the KB were dynamically transported outward to their current orbits after formation closer to the Sun (Morbidelli, 2006).

A key dynamical characteristic of the KB is that many of the orbits there are dynamically 'hot', that is, they have substantial eccentricities and inclinations as high as 40° . These orbits provide strong forensic evidence of one or more dynamical excitation mechanisms that operated in the past. Suggested excitation mechanisms include but are not limited to: scattering by the giant planets, migration of the giant planets, perturbations from formerly present planets in the KB, and external perturbations from passing stars. Hot orbits are also common to the SKB. Most discovered SKBOs were very likely excited to their current orbits by scattering off Neptune.

The CKB also contains a population of bodies on cold (low-eccentricity, low-inclination) nonresonant

orbits (Brown, 2001; Levison and Stern, 2001) that imply formation *in situ*, that is, without the strong dynamical transport of the hot population. This population includes large KBOs, providing strong evidence that there was indeed sufficient mass in the primordial disk extension beyond Neptune for such objects to form.

A cold SKB population has also been proposed (Stern, 1996; Stern and Colwell, 1997a) in order to explain the surface mass density deficit of the KB and the possible *in situ* growth of Sedna (Stern, 2005). Because no such bodies have yet been detected, it is not yet possible to determine definitively whether the primordial protoplanetary disk was compact (i.e., \sim 20–30 AU in radius), a less compact structure (perhaps 50 AU in radius), or a broad disk (which might reach out to 500 AU or more, as observed in many extrasolar debris disks).

10.16.1.3 Kuiper Belt Objects

The properties of KBOs are discussed in greater detail below, but a few relevant facts are as follows. The surfaces of KBOs are relatively unreflective, typically scattering only 3% to perhaps 15% of the light that falls on them, though Eris and Pluto have very high surface albedos (e.g., Grundy *et al.*, 2005). KBOs have a wide range of surface colors (Luu and Jewitt, 1996; Davies, 2001; Jewitt and Luu, 1998, Tegler and Romanishin, 1998; Barucci *et al.*, 1999), varying from almost gray to very red, but it is not clear whether this is due to genetic differences among KBOs or evolutionary effects (e.g., space weathering, collisional resurfacing). It is thought that KBOs consist primarily of mixtures of water ice and rock, with some amount of organic and other complex compounds as well. Most KBOs rotate on their axes in a few hours, but some take days to rotate. Some have sufficiently low enough densities ($<1 \text{ g cm}^{-3}$) to clearly indicate internal macroporosity (e.g., McKinnon, 2002; McKinnon and Barr, 2006). There is already evidence for water and more exotic ices like N₂ and CH₄ on KBOs (e.g., Delsanti and Jewitt, 2006; Licandro *et al.*, 2006). But it is not known if KBOs fall into clear compositional groups as the asteroids do, or whether they instead simply form a broad continuum of surface colors and compositions. Other than Pluto, no KBO is known to have an atmosphere, but at last there are good reasons to expect that some large KBOs exhibit atmospheres, at least transiently (Stern and Trafton, in press).

In mid-2005, it was also discovered that Pluto has two smaller moons (Weaver *et al.*, 2006), making the system a quadruple. Pluto's smaller satellites orbit in nearly circular orbits and in the same plane as Charon; they lie near or at the location of MMRs with Charon. These attributes strongly imply a common origin for all of Pluto's satellites in the giant collision that created Charon (Stern *et al.*, 2006).

In 2001, the first satellite orbiting a KBO, 1998 WW31, was discovered (Velttel *et al.*, 2002). Since that time, numerous other KBOs have been discovered to have satellites. Based on detection statistics and taking into account selection effects that bias detection frequency, it has recently been estimated that at least 10% of all KBOs have satellites (Noll *et al.*, 2002; Margot *et al.*, 2004; Stephens and Noll, 2005); owing to some strong detection biases, the actual fraction of KBOs with satellites may be considerably higher. By late 2006, within 5 years after the discovery of the first KBO binary, there were over two dozen KBOs and centaurs with known satellites. No KBOs except Pluto (with three known satellites) and 2003 EL61 (with two known satellites) are as yet known to have more than one satellite.

Knowledge about KBO satellite albedos is nonexistent; so are colors, except for Pluto's three satellites (Stern *et al.*, 2006). As to satellite spectra, the only published information we are aware of as of mid-2006 is that the brightest satellite of 2003 EL61 exhibits strong water ice features at 1.5 and 2.0 μm that are much deeper than typically found on KBOs (Barkume *et al.*, 2006). We hope that much more information will soon be obtained on KBO satellites, since interesting questions revolve around whether KBOs and their satellites are similar in color, composition, and albedo, or not. Whether the highly dichotomous colors, surface compositions, and albedos of Pluto and Charon are typical, or not, is an open question.

10.16.1.4 On KBO Nomenclature

We, like many others, refer to all bodies orbiting in the KB as KBOs. Therefore, in effect, the KBO designator is merely an indication of the location of bodies orbiting there, rather than an indication of the kinds of bodies orbiting there. The term 'trans-Neptunian object' (TNO) also is used in the literature, and we occasionally use it in what follows, also as an indication of location without any information content about the kind of body that object in this region is.

It is also useful to classify KBOs into size categories as follows:

- *Debris.* KBOs with diameters $D < 50$ m.
- *Comets.* KBOs with $50 \text{ m} < D < 500$ km.
- *Dwarf planets.* KBOs with $500 \text{ km} < D < 5000$ km.
- *Satellites.* Secondary bodies orbiting any KBO.

If KB bodies with diameters >5000 km are later discovered (the SKB region is the most likely region in this regard), we would refer to these objects as planets.

In what follows we will also use the term ‘binary’ from time to time. When we do so, we are referring to two objects that orbit a common center of mass and which have a barycenter in free space between the two orbiting bodies. An example of a KBO binary is thus the binary planet Pluto–Charon.

10.16.2 Environment

The salient environmental characteristics of the KB region are its low population and surface mass density, its cold thermal conditions, and its location within the solar heliosphere. We now discuss the various environmental factors that influence KBOs; these include insolation, collisions, ISM grain erosion effects, and radiation fluence. We describe each in turn. These various influences affect the surfaces of KBOs and therefore shed light on the interpretation of KBO physical measurements, which are discussed later in this chapter.

10.16.2.1 Insolation

The thermal environment of the KB is dominated by the Sun. Although the extant insolation is diluted relative to conditions at 1 AU by factors of order 10^3 – 10^4 , it is still capable of heating surfaces to characteristic temperatures of ~ 20 to ~ 70 K, depending on heliocentric distance, polar obliquity, surface albedo, and surface emissivity.

Because many KB bodies range over great swaths of heliocentric distance (as Pluto does), and have substantial polar obliquities (as Pluto also does), significant annual and seasonal variation in surface temperature are expected on individual KBOs.

At the cold temperatures characteristic of the KB, virtually all cosmogonically abundant molecules have negligible vapor pressures; this includes the common outer solar system ices H₂O and CO₂. The only thermophysically active molecules of cosmogenic importance in the KB temperature regime

are, in order of decreasing volatility with temperature, O₂, N₂, CO, and CH₄. In addition to these four molecular species, the noble gases He, Ne, Ar, and Kr are also volatile in the KB thermal regime. For those KBOs with surface-lying or near-surface-lying volatiles, one expects insolation to drive sublimation, followed by the formation of either global or patchy atmospheres (Stern and Trafton, *in press*).

10.16.2.2 Collisions

As discussed above, collisional processes are important in the KB; this has been known for some time (e.g., Stern, 1992, 1995; see Farinella *et al.* (2000) for a thorough review). Although, owing to the more dilute population density and lower orbital speeds, intrinsic collision rates (the number of collisions per unit surface area per unit time) are lower by a factor of about 1000 in the KB compared to the main asteroid belt, the population of objects there is roughly 1000 times as great. As a result, the overall level of collisional processing of individual objects is of similar scale in the KB as in the main asteroid belt.

Durda and Stern (2000) showed that in the current KB, surface modification by small impactors is expected to be extensive. In particular, they estimated that the cumulative fraction of the surface area of KBOs ranges from a few to a few tens of percent over 3.5 Gy. Such collisions structurally modify surfaces and excavate buried materials, potentially including volatiles, from depths up to several times the impactor diameter.

At typical KB collision speeds, typically 0.5–1.5 km s⁻¹, little surface vaporization of H₂O and CO₂ is expected. What about vaporization of more volatile surface ices? Unfortunately, no studies of the vaporization fraction of more volatiles ices like N₂, CO, and CH₄ are available, so it is not known whether significant mobilization or even loss of these volatiles occurs (E. Pierrazzo, personal communication, 2003). However, by excavating fresh material, collisions may be responsible for creating both high-albedo ejecta rays that influence KBO albedos, and the competition between the excavation of fresh material and its subsequent reddening by radiation may also create surface color variability with time (e.g., Jewitt, 2002; Stern, 2003).

KB collisions happen at sufficiently high speeds that they are erosional for bodies with diameters of a few hundred kilometers and smaller, causing KBO comets and smaller debris to lose surface material over time. In fact, the time-averaged rate of erosion

in the KB is so high that virtually all KBOs with diameters of 5 km and smaller are thought to be fragments ‘chipped’ off larger KBOs in the last 10–20% of the age of the solar system (e.g., Stern, 1996; Davis and Farinella, 1997; Durda and Stern, 2000). Hence, most KB comets are expected to exhibit low surface exposure ages, and are therefore not expected to exhibit many impact craters.

10.16.2.3 Erosional Interactions with the Interstellar Medium

The Sun carries the KB through the interstellar medium (ISM) at a high velocity of approximately 30 km s^{-1} . The resulting impacts of KBOs with ISM grain are important in several ways. They are a contributor of darkening agents to KBO surfaces. They can also erode KBO surfaces, particularly on bodies too small (i.e., with $<500 \text{ km}$ diameter) for the resulting ejecta to remain bound to the target. This process may in total have caused KBOs with sizes of 10 to perhaps 200 km to have lost from about 1 m to perhaps 20 m of surface material since their formation (Stern, 2003).

10.16.2.4 Radiation Processes

The radiation environment of the KB naturally divides into two subcategories: energetic photon bombardment and charged particle bombardment. Interstellar and solar ultraviolet (UV; i.e., $h\nu > 3 \text{ eV}$) photons have copious fluxes in the KB, providing the energy necessary to break bonds and initiate substantial chemical change in Oort Cloud comet surfaces. Because of their much closer proximity to the Sun, KBOs experience a much (i.e., $\sim 10^5$ times) higher UV and solar cosmic ray (SCR) surface doses, dramatically increasing the total deposited charged-particle energy incident on the surfaces of these bodies, relative to Oort Cloud comets. However, the galactic cosmic ray (GCR) dose on KBOs inside roughly 100 AU is expected to be far reduced, relative to Oort Cloud comets, because of the shielding effects of the heliosphere.

UV photo-sputtering is capable of eroding away the uppermost few micrometers of icy surfaces of atmosphereless bodies (Throop, 2000). But more importantly, in a classic series of laboratory experiments and theoretical studies, UV photons would produce significant alteration of the composition, color, and volatility of the upper several to few tens of micrometers of cometary surfaces (e.g., Greenberg,

1982). Thompson *et al.* (1987) and Hudson and Moore (2001) confirmed and extended these results, showing that UV photons promote surface darkening (to albedos of only a few percent) and a devolatilization that becomes progressively more severe with dosage, and therefore age.

Like UV photons, charged particle radiation is capable of both sputtering surfaces and breaking bonds, thereby inducing chemical reactions and consequently reordering the surface ice matrix (e.g., Gerakines *et al.*, 2001). The irreversible radiation-driven conversion of the original, water-dominated ice matrix to a more complex ‘crust’ inevitably leads to the darkening of surfaces, via the formation of longer-chain molecular species, notably hydrocarbons and nitriles. The cosmic ray damage layer on KBOs may reach several meters in depth, depending on the near-surface density and porosity (e.g., Strazzulla and Johnson, 1991).

The red color of Pluto’s surface is widely believed to be the result of radiolysis on its $\text{CH}_4\text{--N}_2\text{--CO}$ surface. The red colors of many KBOs (see below) may also be the result of radiolysis, though there may also be intrinsically red materials on these bodies as well.

10.16.3 Physical Properties

In this section, we summarize the current state of knowledge about the physical properties of KBOs. We begin with their albedos, light curves, and colors. We then discuss KBO surface compositions and atmospheres. We then discuss KBO densities and internal structure. In each area discussed, we compare what is known about the Pluto system to other KBOs.

10.16.3.1 Albedos

Owing to their small sizes on the sky and photometric faintness, few KBOs have measured albedos. The ‘albedo’ of an object is a measure of its reflectivity (the ratio of reflected to incident light). Since the albedo can be a function of wavelength, it is common to refer to the wavelength band or magnitude region of the measurement (e.g., the ‘V-magnitude albedo’ refers to the reflectivity of the object in the wavelength region covered by the commonly used astronomical V (green–red) filter).

Not surprisingly, Pluto and Charon’s albedos are best known. As discussed by Tholen and Buie (1997),

Pluto itself has a V-magnitude albedo of 0.55; Charon's V-albedo is near 0.37.

Initially, many workers assumed that KBOs would have comet-like albedos of 0.04. And indeed, when the first measurements of KBO sizes and albedos were obtained, for two relatively large KBOs (1993 SC and 1996 TL66), using infrared (IR) measurements from the ESA Infrared Space Observatory (ISO), results gave radii of 164 km, corresponding to albedos of 0.02–0.03 (Thomas *et al.*, 2000). Soon, however, other KBOs and centaurs were observed, and were found with much higher albedos; the centaurs Chiron and Asbolus were found to exhibit albedos of 0.17 ± 0.02 and 0.12 ± 0.03 , respectively (Fernandez *et al.*, 2002).

Circumstantial evidence for higher albedos also came from KBO satellite formation studies (Stern, 2003), which indicated the need for typical albedos of 0.15 or higher in order to make collisions sufficiently common to explain the growing population of KBO binaries. Then KBO 2002 AW197 was found to have an albedo of 0.17 ± 0.03 (Cruikshank *et al.*, 2005), and Lykawka and Mukai (2005) and Grundy *et al.* (2005) demonstrated a tendency for increasingly larger objects to have increasingly higher albedos (**Figure 1**).

Table 1 reveals that few measured KBO albedos are <5%, and few are higher than 20%; instead, most measured albedos lie in the range $15 \pm 7\%$. The few KBOs with very high albedos, like Pluto and Eris, may – like Pluto – exhibit atmospheric recycling

mechanisms that refresh the surface albedo (e.g., Stern *et al.*, 1988).

It will be interesting to see how increasingly larger albedo data sets affect these conclusions, and whether distinctions appear with dynamical class.

10.16.3.2 Rotational Properties and Light Curves

Rotation rates and light curves of KBOs hold clues to their shapes and surface variegation, as well as to their rotational dynamics. The definitive detection of Pluto's rotation rate came from photoelectrically observing its 6.4-day rotational light curve over 50 years ago (Walker and Hardie, 1955). Recently, it has been possible to use HST to separate the individual light from Pluto and Charon (Buie *et al.*, 1997; see **Figure 2**), showing rotation rates and orbital period of the pair are synchronous, which is consistent with expectations from tidal evolution theory (e.g., Dobrovolskis *et al.*, 1997). Of more direct importance to our interests here, Charon's low-amplitude (~8% peak to peak) light curve was shown to exhibit little variation, quite unlike Pluto's high-amplitude (~30% peak to peak) light curve.

Over 65 KBOs and centaurs now have measured light curves or light curve upper limits (Trilling and Bernstein, 2006), as shown in **Figure 3**. Approximately half of these have established light curve periods and amplitudes. All measurements of KB light curve periods except Pluto's lie between 3.5

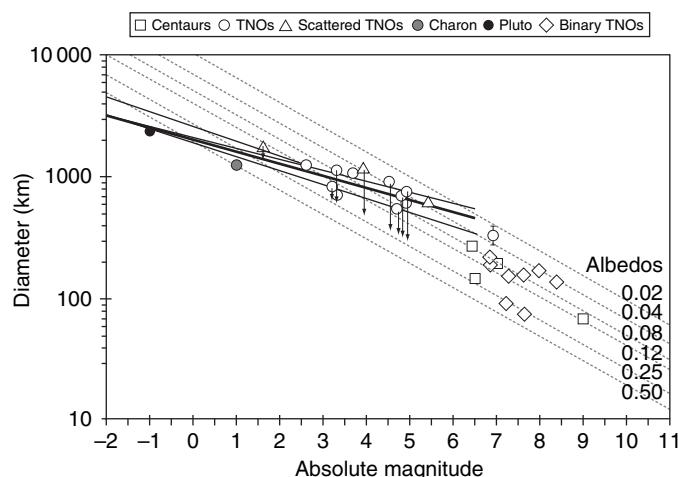


Figure 1 Some KBO (aka TNO) and Centaur albedos determined from measured sizes and absolute magnitudes. The dashed line shows various albedo values that can be assumed when size measurements are not available. From Lykawka PS and Mukai T (2005) Higher albedos and size distribution of large trans-Neptunian objects. *Planetary and Space Science* 53: 1319–1330.

Table 1 Measured V- and R-magnitude albedos of KBOs

Object	Dynamical class	Albedo (%)	Method	Reference
(58534) 1997 CQ29	Classical	39 ± 17	Binary orbit	G2005
(19521) Chaos	Classical	≥ 5.8	Thermal (1.2 mm)	G2005
1998 WW31	Classical	6.0 ± 2.6	Binary orbit	G2005
(66652) 1999 RZ253	Classical	29 ± 12	Binary orbit	G2005
(20000) Varuna	Classical	$3.7^{+1.1}_{-1.4}$	Thermal (0.9, 1.2 mm)	G2005
(88611) 2001 QT297	Classical	10 ± 4	Binary orbit	G2005
(50000) Quaoar	Classical	10 ± 3	Direct imaging	G2005
2003 EL61	Classical	≥ 60	Binary orbit	R2006
(26308) 1998 SM165	2:1	14 ± 6	Binary orbit	G2005
(15789) 1993 SC	3:2	$3.5^{+1.6}_{-1.3}$	Thermal (90 μm)	G2005
(19308) 1996 TO66	3:2	≥ 3.3	Thermal (1.2 mm)	G2005
(47171) 1999 TC36	3:2	22 ± 10 $7.9^{+3.1}_{-2.4}$	Binary orbit Thermal (24, 70 μm)	G2005 S2006
(28978) Ixion	3:2	≥ 14.8	Thermal (1.2 mm)	G2005
(24835) 1995 SM55	Scattered	≥ 6.7	Thermal (1.2 mm)	G2005
(15874) 1996 TL66	Scattered	≥ 1.8	Thermal (90 μm)	G2005
(38628) Huya	Scattered	≥ 8.4	Thermal (1.2 mm)	G2005
2001 QC298	Scattered	2.5 ± 1.1	Binary orbit	G2005
(55565) 2002 AW197	Scattered	$10.1^{+3.8}_{-2.2}$ 17 ± 3	Thermal (1.2 mm) Thermal (24, 70 μm)	G2005 C2005
(55636) 2002 TX300	Scattered	≥ 19	Thermal (1.2 mm)	G2005
(84522) 2002 TC302	Scattered	5.1	Thermal (1.2 mm)	G2005
(90377) Sedna	Scattered	4.6	Thermal and imaging	G2005

C2005: Cruikshank *et al.* (2005) V-band; G2005: Grundy *et al.* (2005) R-band; R2006: Rabinowitz *et al.* (2006); S2006: Stansberry *et al.* (2006) V-band.

and 27 h. Only three exceed 16 h. Hence, to first order, KBO light curve periods appear to be similar to asteroid light curve periods. Most measured KBO light curve amplitudes are <25%. It is not known what relative roles shape effects and albedo effects play in the observed population of measured light curve amplitudes.

10.16.3.3 Colors

The faintness of KBOs makes it difficult to measure accurate photometry and colors. This is problematic in several respects, including the biasing of available measurements to brighter, and therefore typically larger, KBOs.

As with rotation albedos and periods, the first color information obtained on objects in the KB was from measurements of Pluto. Pluto's surface was determined to be red as early as the 1950s. Later, color variations were observed with rotational phase, and color maps were made from HST (Stern *et al.*, 1997) and mutual event data (e.g., Binzel, 1988; Young *et al.*, 2001). In contrast to Pluto, Charon's surface appears to be spectrally neutral, with no significant rotational variation.

The first collections of KBO color data were published in the late 1990s (see Luu and Jewitt, 1996; Davies, 2001; Jewitt and Luu, 1998; Tegler and Romanishin, 1998; Barucci *et al.*, 1999). However, whereas Luu and Jewitt (1996) found a continuous range among the colors of their 13 objects (four centaurs and nine KBOs), Tegler and Romanishin (1998) found that KBO colors fell into two distinct populations separated by a color gap: one population had colors slightly redder than the Sun, and the other had colors significantly redder, containing the reddest objects in the solar system.

This disagreement spurred observations by several groups, none of whom found evidence for a bimodal color distribution (e.g., Barucci *et al.*, 2000; Delsanti *et al.*, 2001; Boehnhardt *et al.*, 2001). A statistical, principal components analysis of 22 KBOs (Barucci *et al.*, 2001) also found that the colors of KBOs and centaurs were consistent with a continuous (rather than a bimodal) distribution, and when Jewitt and Luu (2001) analyzed a larger data set than in their 1996 and 1998 papers, they again found a continuous distribution. However, when Tegler and Romanishin (2003) performed a statistical analysis of 107 KBOs (50 objects from their own observations and 57 objects observed by other

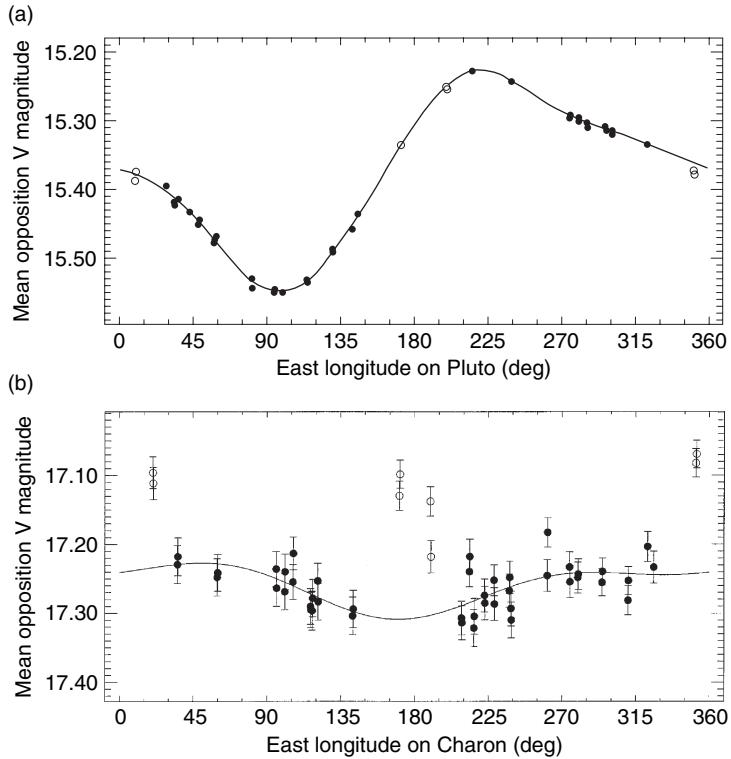


Figure 2 Rotational light curves of Pluto (a) and Charon (b). The mean opposition V magnitude is plotted against sub-Earth longitude in degrees. For Pluto, the error bars are smaller than the symbols and are therefore not shown. The open circles were not included in the fit and correspond to points in the orbit when Pluto and Charon are at a minimum separation. The solid curves represent best-fit light curves. From Buie MW, Tholen DJ, and Wasserman LH (1997) Separate light curves of Pluto and Charon. *Icarus* 125: 233–244.

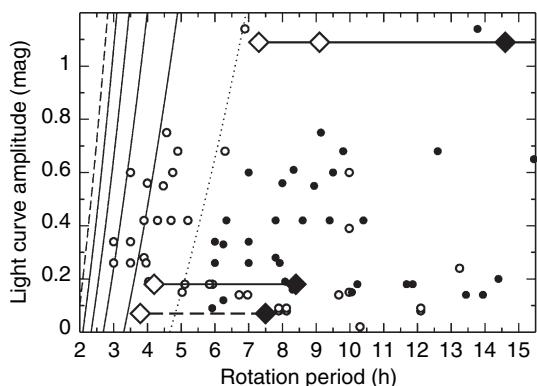


Figure 3 Measured light curve amplitudes of KBOs and Centaurs (Trilling and Bernstein, 2006; TB06). Diamonds represent TB06's own data; circles represent data from other sources. Open symbols indicate photometric periods, and filled symbols indicate rotation periods that are twice the observed photometric period. Objects known to have double-peaked light curves are plotted only as filled symbols. The curves represent solutions for critical rotation periods for various densities in grams per cubic centimeter: 0.5 (dotted line), 1, 1.5, 2, 2.5, and 3 (dashed line). Regions above and to the left of a given curve are dynamically unstable for a given density.

groups), they again found a bimodal distribution, which they argue is not detectable in the measurements of other groups due to larger average photometric uncertainties. Peixinho *et al.* (2003) countered that the bimodal distribution exists within the centaurs, but not KBOs themselves (Figure 4).

Currently, there is a broad consensus that KBOs and centaurs display a wide range of colors; but it is not clear whether the color distribution is continuous or separated into distinct groups, though the majority of papers claim the former.

What is the reason for the wide range of KBO colors? Is it related to their origin (e.g., composition) or their evolution (e.g., owing to impacts, transient atmospheres, and varying radiation environments)? Alterations by impacts can compete with irradiation by cosmic rays, so the color of objects in different regions could vary with the local rate of impacts (Gil-Hutton, 2002).

To attempt to constrain the origin versus evolution scenarios further, a variety of workers have searched for color correlations with various orbital and physical parameters in order to address the

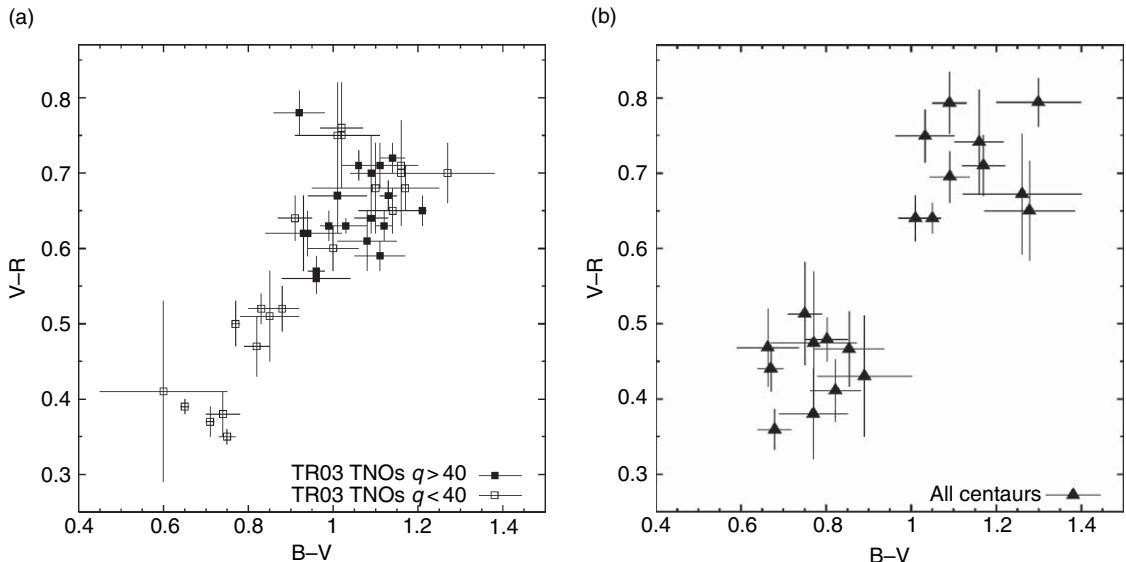


Figure 4 V-R and B-V magnitude color–color plot for KBOs (a) and centaurs (b). B-V is the brightness difference (hence color) seen through the commonly used ‘blue’ and ‘visible’ standard astronomical filters, and V-R is the color between the ‘visible’ and ‘red’ filters. From Peixinho N, Doressoundiram A, Delsanti A, Boehnhardt H, Barucci MA, and Belskaya I (2003) Reopening the TNOs color controversy: Centaurs bimodality and TNOs unimodality. *Astronomy and Astrophysics* 410: L29–L32.

source of the KBO color diversity. Tegler and Romanishin (1998) did not find any correlation of their color populations with orbital parameters of the objects, speculating that other factors in the objects’ history must come into play, such as solar UV photons, solar-wind and solar-flare particles, and cosmic rays. Barucci *et al.* (1999) discussed the compositional surface variety among Centaurs, but also found no relationship with orbital parameters, for example, perihelion distance. Tegler and Romanishin (2000) however reported that all nine KBOs in their survey with nearly circular orbits with perihelion distances greater than 40 AU have very red surfaces. This work was later extended to show that the color variations with different dynamical classes point toward a primordial origin, with gray objects forming closer to the Sun than red objects. Hainaut and Delsanti (2002) found that cold classical belt objects apparently have significantly redder colors, whereas the plutinos have the opposite trend. However, Trujillo and Brown (2002) found a correlation between the BVR colors and the orbital inclination of classical KBOs. McBride *et al.* (2003) found a weak correlation between inclination and color in the classical KB, which they more specifically characterize as a lack of blue objects with low inclinations. Stern (2002) found a correlation

between colors of KBOs and the mean random impact speed the objects would experience as a function of their orbital eccentricities and inclinations. This is consistent with KBO colors being driven by a competition between collisional excavation of neutral material and subsequent radiation reddening. Thébault and Doressoundiram (2003) and Doressoundiram *et al.* (2005) also found a correlation of color with mean impact velocity.

At present, the root cause or causes of the color variation among KBOs remain ambiguous. There is clearly no shortage of ideas; instead, however, there is a shortage of data from which to make strong conclusions. It may require several hundred or several thousand KBOs with well-established colors for definitive progress to be made. Progress will also derive from studies of color variegation on KBOs as they rotate, and with the eventual, hoped-for return of color imagery of KBOs by *New Horizons*.

10.16.3.4 Surface Composition

As with other types of physical studies, the spectra of objects in the KB were taken on Pluto. Methane frost was identified on Pluto’s surface by Cruikshank *et al.* (1976). Owen *et al.* (1993) later added the identification of N₂ and CO to the inventory of Pluto surface

frosts. Owen *et al.* (1993) also recognized that methane must be mixed at the molecular level with nitrogen because of the shift in the methane band centers. Later, Douté *et al.* (1999) identified more bands of CO and CH₄ on Pluto's surface, in an analysis of IR spectra from 1.4 to 2.5 μm (Figure 5). Recently, Cruikshank *et al.* (2006) reported the detection of C₂H₆ (ethane) on Pluto's surface.

Studies by Douté *et al.* (1999), Grundy and Fink (1996), and Grundy and Buie (2001) provided important information of the compositional heterogeneity of Pluto's surface (Figure 6). They found that CO

frost is more abundant peak in the sub-Charon hemisphere, and that two distinct units containing CH₄ frost exist: one described by the strong methane bands that correlates with the visible light curve, and another described by the weak methane bands that does not correlate with the visible light curve.

Pluto's surface variegation in albedo, color, and composition is likely due to a combination of internal activity, volatile transport, and atmosphere–surface interactions; impacts may also play a role. At present, the available data sets do not have sufficient

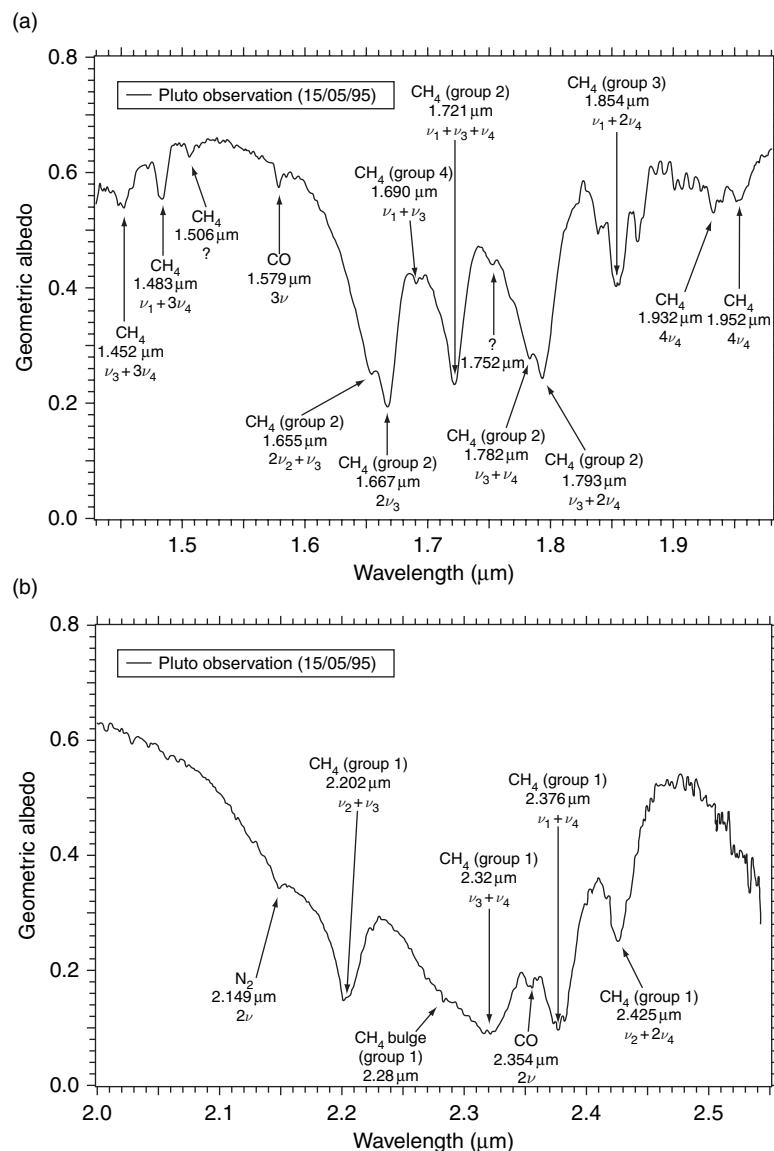


Figure 5 Pluto's spectrum and identification of various absorption bands. From Douté S, Schmitt B, Quirico E, *et al.* (1999) Evidence for methane segregation at the surface of Pluto. *Icarus* 142: 421–444.

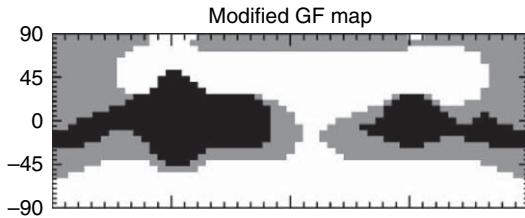


Figure 6 A Pluto composition map in cylindrical projection (Grundy and Buie, 2001). White areas are N_2 -rich terrains, gray areas are CH_4 -rich terrains, and black areas are $\text{H}_2\text{O}/\text{tholin}$ -rich terrains.

resolution to distinguish uniquely between the various possibilities in order to determine which aspects of the variegation at which locations are due to specific processes.

Owing to Charon's close angular proximity to Pluto, spectroscopy of Charon itself was first possible only as a result of mutual event observations. Such spectra were obtained in the visible by Fink and DiSanti (1988) and in the near-IR by Marcialis *et al.* (1987) and Buie *et al.* (1987). This technique involved subtracting the Pluto-only spectrum (when Charon was behind Pluto) from the combined spectrum taken just adjacent to the eclipse event. The resulting visible-wavelength spectra showed Charon to be neutral in color, but the near-IR spectra of Charon obtained this way detected water ice on Charon's surface.

HST IR observations using the NICMOS IR imaging spectrometer (Buie and Grundy, 2000) were later obtained at four different longitudes in 1998. These spectra revealed that the near-IR spectrum of Charon varies only weakly with longitude, and is consistent with crystalline water ice with grain sizes in the tens of micrometers (Brown and Calvin, 1999; Buie and Grundy, 2000), plus an unknown spectrally indistinct absorber. Ground-based IR spectra obtained in very good seeing conditions using the Keck 10 meter (Brown and Calvin, 1999) also revealed evidence for crystalline H_2O -ice and the possible presence of ammonia hydrates, suggesting the possibility of geologic activity. Cook *et al.* (2006) later confirmed these findings.

Spectroscopy of KBOs is considerably more difficult than Pluto and Charon because they are so much fainter. Most KBOs that have been studied show featureless reflectance spectra. However, strong 1.5 and 2.0 μm water ice features were detected on the Centaur (10199) Chariklo by Brown and Koresko (1998), on Pholus by Cruikshank *et al.* (1998), and on Chiron by Foster *et al.* (1999) and later Luu *et al.*

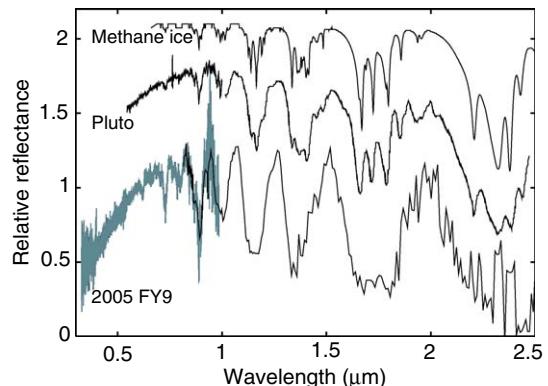


Figure 7 The spectrum of 2005 FY9 compared to Pluto and pure methane ice (Licandro *et al.*, 2006); the latter two spectra were shifted vertically for clarity.

(2000). Water ice features were first detected on a KBO by Brown *et al.* (1999) in a spectrum of 1996 TO66. Interestingly, these data indicated possible variations with rotational phase, perhaps indicating a patchy surface. Licandro *et al.* (2001) found deep water ice absorption in 2000 WR106. IR observations of the KBO (50000) Quaoar detected both crystalline water ice and ammonia hydrate (Jewitt and Luu, 2004). These data were interpreted to indicate that Quaoar has recently been resurfaced either from impacts or cryovolcanic activity. Crystalline water ice was also clearly detected in spectra of 2003 EL61 (Trujillo *et al.*, 2006), again implying a resurfacing process taking place.

As to constituents other than water ice, the available data set is meager. Perhaps the most interesting results are IR spectra of the large KBO 2005 FY9, which appears to be very similar to Pluto, but with even stronger prominent absorption bands of CH_4 (Licandro *et al.*, 2006; see Figure 7). The tentative detection of N_2 on Sedna and Eris and CH_4 on Sedna has also been reported (Licandro *et al.*, 2006).

10.16.3.5 Atmospheres

The only known KBO with an atmosphere is Pluto, which possesses an atmosphere with a base pressure of order $10 \mu\text{b}$ and a composition that is largely N_2 , with traces of CO and CH_4 . For excellent reviews, see Trafton *et al.* (1997) and Yelle and Elliott (1997).

Perhaps the most important aspect of Pluto's atmosphere for the purpose of this review is that it seasonally collapses to the surface as Pluto draws away from the Sun and cools (e.g., Trafton and Stern, 1983; Stern and Trafton, 1984). This results

in the seasonal/orbital transport and deposition of a high albedo frost of thickness some millimeters to centimeters that persists for most of Pluto's orbit, when the atmosphere cannot be supported by thermally driven ice sublimation (Stern *et al.*, 1988).

In order to support a significant atmosphere, a KBO must either exhibit internal outgassing or possess volatile ices that are warmed sufficiently to generate a non-negligible vapor pressure equilibrium, or both. Indeed, with the discovery of CH₄ frost on Pluto's surface (Cruikshank *et al.*, 1976), the possibility of an atmosphere supported by vapor pressure equilibrium was already inferred.

This same process will occur on KBOs that generate seasonally or orbitally forced atmospheres, if the KBO is large enough to retain its atmosphere against immediate hydrodynamic (i.e., comet-like) escape. The estimated diameter for a KBO to retain its bound atmosphere is about 700 km (Lykawka and Mukai, 2005). For KBOs in excess of this size that possess atmospheres, one would expect a cyclic 'atmospheric laundry' to renew the surface albedo, as at Pluto (Stern *et al.*, 1988).

10.16.3.5.1 Pluto's atmosphere

Since vapor pressures have an exponential dependence in temperature, frost temperature is a critical parameter for understanding the density and pressure of Pluto's atmosphere. Temperature observations of Pluto's surface first measured the flux at the IR and millimeter wavelengths. Altenhoff *et al.* (1988), Stern *et al.* (1993), and Jewitt (1994) found brightness temperatures of $\sim 30\text{--}40$ K from millimeter measurements at 800–1200 μm . A similar result was obtained using the temperature-diagnostic 2.148 μm feature of N₂ as a thermometer (Tryka *et al.*, 1994). However, Infrared Astronomical Satellite (IRAS) measurements by Sykes *et al.* (1987) found the brightness temperature to be warmer (55 K) at 60 μm , near the peak of Pluto's black body emission. The discrepancy between the IRAS and larger wave temperature measurements was and still is attributed to the presence of both hot (inert) and cold (sublimating) regions on Pluto's surface. Stern *et al.* (1993) predicted that if the latter was in fact the case, Pluto's visible and thermal light curves should be inversely correlated. Subsequently, observations using the ISO spacecraft detected a thermal light curve for the Pluto–Charon system at 60 μm , indicating that Pluto's surface is not isothermal (Lellouch *et al.*, 2000), with maximum dayside temperatures of 54–63 K.

Combining information from surface-temperature measurements and surface spectroscopy, one infers that either N₂ or CO must dominate Pluto's atmosphere, and that CH₄ must be present only in trace amounts. N₂ is expected to dominate CO based on Raoult's law, that is, because it is dominant on the surface and the two have similar volatilities.

It was not until 1988 that Pluto's atmosphere was definitively detected using the technique of stellar occultation (Hubbard *et al.*, 1988; Elliot *et al.*, 1989). All of the 1988 stellar occultation light curves exhibit a sharp change of slope, or 'knee' during immersion and emersion (see Figure 8).

The structure of Pluto's atmosphere as probed by the 1988 occultation was well modeled by an isothermal upper atmosphere (above a radius of 1215 km) with a temperature of 104 ± 2 K for composition of nitrogen. This temperature is significantly greater than any surface temperature measurements for Pluto. The 'knee' in the occultation light curve can be explained by either a temperature gradient of 10 to 30 K km^{-1} (Eshelman, 1989; Hubbard *et al.*, 1990; Stansberry *et al.*, 1994) or a haze layer (Elliot *et al.*, 1989). These mechanisms are not mutually exclusive and both may be or have been at work.

Yelle and Lunine (1989) realized that CH₄ can act as a thermostat in Pluto's upper atmosphere by absorbing energy at the v_3 band at 3.3 μm , cooling via the v_4 band at 7.6 μm , and conducting heat to the surface. This mechanism may explain the high temperatures at the microbar pressure level in light of the colder surface temperatures. Strobel *et al.* (1996) expanded on this model and developed a radiative-conducting atmospheric model, including the effects

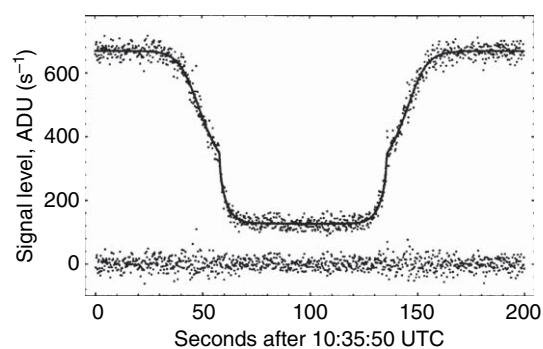


Figure 8 Data (points; Elliot *et al.*, 1989) and model (line; Elliot and Young, 1992) of the 1988 stellar occultation light curve of Pluto observed with NASA's Kuiper Airborne Observatory. The points near the zero level along the bottom are the residuals from the fit of the model to the data.

of heating and cooling by methane at 3.3, 2.3, and 7.6 μm , cooling by CO rotational line emission (originally considered by Lellouch *et al.* (2000)), and heat conduction. Doing so, they found that the 2.3 μm band was more important for heating the atmosphere than the 3.3 μm band. Their preferred model can reproduce the temperature gradient of the occultation light curve, but has an upper atmosphere temperature that is too warm (134 K).

Of course, all of the ‘ CH_4 -thermostat’ models for Pluto’s atmospheric thermal structure rely on sufficient quantities of CH_4 in the atmosphere. And while CH_4 had been inferred indirectly from the combination of its known presence in solid state on the surface and surface thermal measurements, its actual atmospheric abundance was highly uncertain because it was not known if the CH_4 source regions were in Pluto’s cold or hot regions.

Spectroscopic detection of gaseous CH_4 was first obtained in 1994 (Young *et al.*, 1997) with a subsequent confirmation in 2006 (Roe, 2006). This work confirmed that the gaseous CH_4 in Pluto’s atmosphere is only a trace constituent. The partial pressure of methane is overabundant for methane in solid solution with nitrogen, but consistent with methane in vapor-pressure equilibrium with pure CH_4 at 41–45 K, as shown in Figure 9.

From the well-observed stellar occultation of August 2002, it was discovered that Pluto’s atmospheric pressure had doubled since 1988 (Elliot *et al.*, 2003; Sicardy *et al.*, 2003; Figure 10). The increase in atmospheric pressure is consistent with a change in nitrogen frost temperature by 1.3 K (Elliot *et al.*, 2003), which may be due to a thermal phase lag following the 1989 perihelion maximum in insolation (e.g., Stern *et al.*, 1988; Trafton, 1990; see also the review by Spencer *et al.* (1997)).

In addition to the increase in atmospheric pressure, an increase of turbulence in Pluto’s atmosphere was evidenced by more refractive spikes in the occultation light curves of 2002, compared with 1988. These spikes indicate the presence of waves or turbulence and were practically nonexistent in the 1988 observations. These spikes were also seen in the lower atmosphere in the 2002 observations (Pasachoff *et al.*, 2005).

Another important change seen from 1988 to 2002 is the declining strength of the kink in the occultation light curves, relative to what was seen in 1988. This indicates that a change in the atmosphere’s vertical structure has also taken place. Elliot *et al.* (2003) reported on observations of the August 2002 occultation at effective wavelengths from 0.75 to 2.2 μm . The minimum flux of the light curve varies with

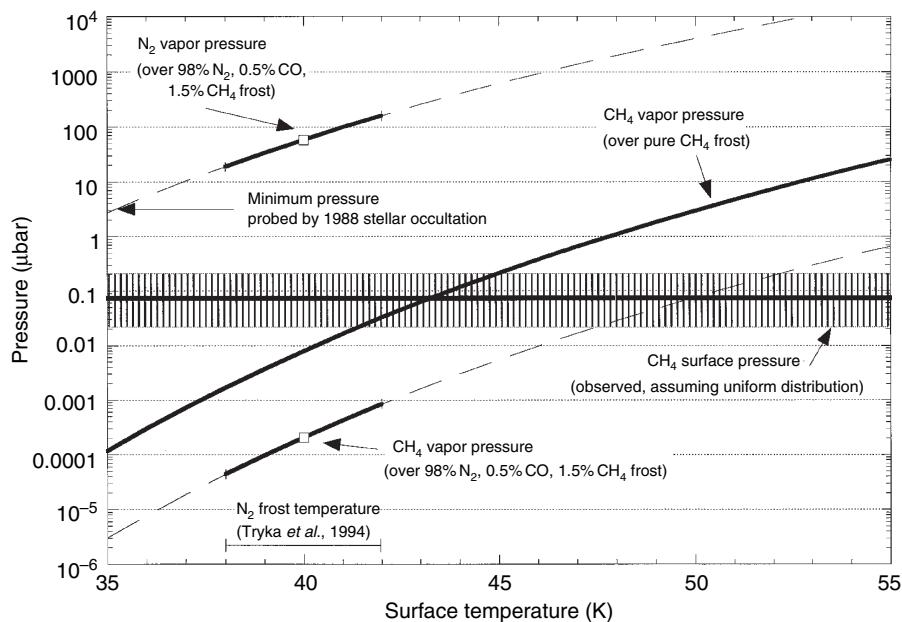


Figure 9 Surface vapor pressures on Pluto as a function of temperature of the sublimating material. The solid parts of the various curves show the vapor pressures for a frost temperature of 40 ± 2 K. From Young LA, Elliot JL, Tokunaga A, de Bergh C, and Owen T (1997) Detection of gaseous methane on Pluto. *Icarus* 127: 258–270.

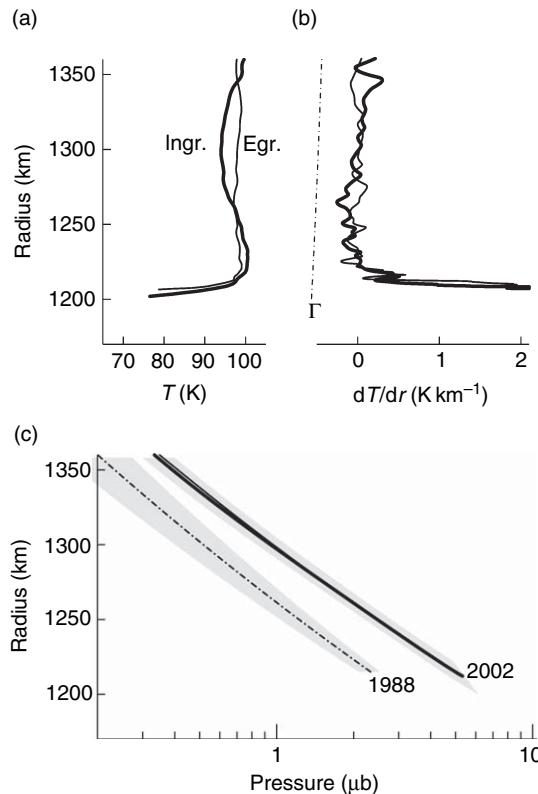


Figure 10 Temperature and pressure profiles of Pluto's atmosphere derived by Sicardy *et al.* (2003) from the inversion of the occultation light curve, assuming a spherically symmetric and transparent pure nitrogen atmosphere. (a) Pluto's atmospheric temperature profiles $T(r)$ at ingress (thick curve) and egress (thin curve). The radius is the distance to the planet center. (b) The corresponding temperature gradient profiles dT/dr . The dash-dotted line provides the adiabatic lapse rate profile $\Gamma = -g/c_p$, to the left of which the atmosphere would become convectively unstable, where $c_p = 103 \text{ J K}^{-1} \text{ kg}^{-3}$ is the specific heat at constant pressure for N_2 . (c) The corresponding pressure profiles. The dash-dotted line is the pressure profile derived from the Kuiper Airborne Observatory light curve of 9 June 1988, with the associated error domain as the shaded region. The ingress and egress profiles derived from the Canada-France-Hawaii Telescope (CFHT) light curve of 2002 are the nearly coincident thick and thin solid lines, respectively.

wavelength, indicating that there is extinction in Pluto's lower atmosphere, that is, a haze layer. A purely refracting atmosphere would produce a light curve with the same minimum flux regardless of the wavelength.

All of the effects seen in the 2002 occultations were seen again in 2006, but the number of turbulence-driven refractory spikes appears to have further increased (e.g., Young *et al.*, 2006; Elliot

et al., 2006; Sicardy *et al.*, 2006). Clearly, Pluto's atmosphere is changing, most likely due to post-perihelion changes in insolation and subsolar latitude. When will atmospheric collapse take place? We do not know, but models (e.g., Trafton, 1990; Trafton *et al.*, 1997) indicate it will occur before 2020.

10.16.3.5.2 Other KBOs

Regarding centaurs, 2060 Chiron has long been known to exhibit both photometric variability (Hartmann *et al.*, 1989) and a cometary coma (Meech and Belton, 1989). Somewhat surprisingly, Chiron has exhibited activity at all points of its orbit with no clear correlation between the level of cometary activity and heliocentric distance. In fact, Chiron was more active at its aphelion in 1970 than it was in following decades when it was closer to the Sun. The level of activity can vary by more than a factor of 2 and last anywhere from weeks to months. Bus *et al.* (1991) detected CN in Chiron's coma, and Womack and Stern (1999) reported a detection of CO, which may be the sublimation agent responsible for Chiron's cometary activity. This suggestion is consistent with the models indicating that Chiron's activity can be explained by a composition of CO near the surface as an ice or as a gas trapped in the amorphous ice that is released by delayed thermal pulses (Prialnik *et al.*, 1995; Capria *et al.*, 2000). The variability of Chiron's outbursts, the apparent lack of a correlation between activity and distance from the Sun, and the observations of discrete jet-like features observed during an occultation (Elliot *et al.*, 1995) imply that Chiron's surface may be unevenly covered with volatile material, and may exhibit geysers like those seen on Triton (e.g., Smith *et al.*, 1989).

In contrast to the known atmosphere around Pluto, only upper limits have been placed on possible atmospheres around other KB bodies.

Although no atmospheres have as yet (November 2006) been detected around KBOs other than Pluto, Hainaut *et al.* (2000) found a change in the light curve of KBO 1996 TO66 from a double peak to a single peak over the course of a year, which they suggested may have resulted from an episode of cometary activity. Further, the discoveries of N_2 , CO, CH_4 , and C_2H_6 (ethane) on the KBO 2005 FY9 (Brown *et al.*, 2006; Licandro *et al.*, 2006), and the tentative detection of N_2 on Sedna and Eris and CH_4 on Sedna (Licandro *et al.*, 2006), indicate that KBO atmospheres may be discovered in the future. However, given the exponentially sensitive dependence of vapor pressures to temperature, and therefore to

heliocentric distance and seasonal (obliquity-driven) effects, it is natural to predict that such atmospheres should in many cases be transient with seasonal or orbitally driven cycles, as Pluto's is.

Finally, it is worth noting that the captured dwarf planet Triton, now a satellite of Neptune, is known to possess an N₂-dominated atmosphere with some CH₄ and a base pressure not unlike Pluto's. Since Triton is also roughly Pluto's size and density, and is widely thought to have originated in the same region of the solar system as Pluto, it serves as a possible guide to future discoveries that may be made around dwarf planets in the KB and farther out as well.

10.16.3.6 Densities

With the discovery of two new satellites in the Pluto system (Weaver *et al.*, 2006), Buie *et al.* (2006) have been able to derive a better estimate of the mass ratio of Charon to Pluto, and therefore the individual masses of each, than had previously been possible. Combining this new mass ratio with the recently improved 604 ± 3 km radius of Charon from the stellar occultation in 2005 (Sicardy *et al.*, 2006), we now also have a much more accurate estimate for Charon's density, 1.66 ± 0.06 g cm⁻³ (Buie *et al.*, 2006). The current best estimate of Pluto's density is 2.03 ± 0.06 g cm⁻³ (Buie *et al.*, 2006).

Densities for KBOs are only available for a subset of the two dozen objects with satellites. In such binary systems, we can determine the mass ratio of the satellite to the primary from the KBO–satellite mutual orbit, but we also need an estimate of the diameter to derive the density; the latter requires direct resolution of the KBO's size by HST, thermal IR measurements, or mutual events – all of which are difficult to obtain. As a result, only a few KBO and centaur densities are available. Spencer *et al.* (2006) reported a density of $0.51(+0.29, -0.11)$ g cm⁻³ for KBO (26308) 1998 SM165; this is comparable to the $0.5\text{--}0.8$ g cm⁻³ density derived for KBO (47171) 1999 TC36 (Stansberry *et al.*, 2006). Grundy *et al.* (2006) reported a density for Centaur (65489) 2003 FX128 of $1.38(+0.65, -0.32)$ g cm⁻³. However, Rabinowitz *et al.* (2006) reported a far higher, model-dependent density of $2.6\text{--}3.3$ g cm⁻³ for 2003 EL61 from studies of this KBO's rotational dynamics.

Clearly the study of KBO densities is in its infancy. Only a handful of results are available. All concern large KBOs and all but one concern binary KBOs. However, already there appears to be a wide diversity in KBO densities, ranging from ~ 0.5 to

>2 g cm⁻³. This in turn implies a wide range of internal porosities and water-ice/rock fractions in the KBO population.

10.16.3.7 Interior Models

Several groups have begun to investigate the thermal evolution of KBOs. One motivation for such studies is the current-day presence of surface volatiles on Pluto after what appears to be 4 Gy of active atmospheric escape. As demonstrated some time ago (Stern, 1989a), this argues for the possibility of present-day internal activity to pipe new volatiles to compensate escape losses. Another motivation for such studies results from the sporadic activity of the 160-km-wide centaur Chiron, which shows no correlation with heliocentric distance, solar activity, or obliquity-derived season (e.g., Bus *et al.*, 1991; Duffard *et al.*, 2002).

Models show that radiogenic heating and solar irradiation of an initially undifferentiated KBO can sublime and deplete most of the volatile ices near the surface. The resulting internal heating can result in a compositionally layered structure, with interlaced layers depleted and enriched in volatiles. When such bodies are injected into the inner solar system, episodic activity can occur as the volatile-rich layers reach sublimation temperatures. Thermal evolution can also be affected by the heating of collisions, which can sublime ices, induce gas flow in porous structures, and eject gas and dust. Depths thermally affected by KBO impacts can vary by large amounts, and to depths greater than 1 km (Orosei *et al.*, 2001).

To better understand interior evolution, McKinnon (2002) and Choi *et al.* (2002) conducted studies of the thermal evolution of KBOs of various sizes up to almost 1000 km in diameter. These workers agree that ²⁶Al cannot be an important heat source, owing to the long accretion timescale for KBOs relative to the 0.73 My half-life of ²⁶Al. They also found that for 100 km KBOs, radiative losses to space prevent central temperatures from exceeding 80 K; for 150 km KBOs, the central temperature limit found was near 105 K. However, for a 450-km-radius KBO, central temperatures can exceed 270 K; this opens the possibility of liquid H₂O in large KBOs and Pluto.

Moreover, McKinnon (2002) found that even KBOs in the 75–225 km radius range can transform H₂O-ice from its amorphous to crystalline state in their interiors. Other important thermal benchmarks include 155 K, where a eutectic CH₃OH–H₂O melt

can form, and 190 K, where solid-state H₂O-ice creep and hence sintering/densification become important. McKinnon (2002) concludes that over a wide range of parameter choices, long-lived radioactive energy release probably powered activity in the largest KBOs for hundreds of millions of years, at least. These findings indicate Pluto and other large KBOs may well show early, and perhaps even persistent, signs of internal activity.

Most recently, McKinnon and Barr (2006) and Hussman *et al.* (2006) have discussed the possibility of internal liquid H₂O oceans in large KBOs, including both Pluto and Charon. For rock ratios of 0.65 (consistent with Pluto's density) and surface temperatures near 40 K, internal oceans were commonly found, though at a range of interior depths and with various thicknesses. This exciting possibility begs for tighter constraints on KBO densities and compositions, so as to better constrain the range of parameter space available to modelers.

10.16.4 Origin

Unlike early ideas in which the trans-Neptunian debris belt contains only leftover material in the form of very small bodies from the era of giant planet formation (e.g., Kuiper, 1951), the real KB actually contains very large numbers (i.e., $>10^5$) of bodies that reached sizes of 100 km, and at least several dozen and more likely many times more objects (depending on the large body population of the SKB) that reached 1000–3000 km in diameter.

The standard paradigm of planet formation involves three major stages. We review this paradigm briefly here, but refer the reader to excellent, succinct reviews with more detail, such as Ward (1996) and *The Origin of the Natural Satellites*.

The first stage of planet formation involves the coagulation of dust grains into macroscopic objects that grow by gentle accretion when physical contact occurs, but at a rate dominated purely by two-body, physical (i.e., projected surface area) collision cross sections. Stage 2, the so-called mid-stage of accretion, begins when collision rates are substantially affected by gravitational cross sections, which accelerates growth from an R² process to an R⁴ process. Mid-stage growth terminates when bodies self-isolate by growing to a size where they have depleted the feedstock of smaller material within their gravitational reach. The third and final stage of accretion results from the mergers of the planetary

'embryos' created in stage 2. These mergers occur on longer timescales driven by dynamical effects that cause the embryos to perturb one another onto mutually crossing orbits that inevitably cause collisions among relatively like-sized bodies.

In order for solid-body accretion to proceed, two key conditions must simultaneously obtain on a relevant timescale. These are (1) a sufficient supply of material to promote collisions on the relevant timescale, and (2) sufficiently gentle collisions to permit growth, rather than erosion and/or destruction.

Shortly after the discovery of the first KBOs, when crude estimates of the population of the KB first became available, it was recognized (Stern, 1995) that the current-day KB does not satisfy either of these two conditions for accretion. More specifically, it was pointed out that (1) the surface mass density of the current day KB is too low by a factor of approximately 10³ to allow the growth of the observed KBOs in the age of the solar system, and (2) current-day collision velocities (calculated from the mean random crossing speeds of intersecting orbits) are too high by a factor of about 10⁴ to initiate and sustain stage 1 growth. These facts in turn led to the conclusion (Stern, 1995, 1996) that the current KB must be highly evolved from its primordial state, in which both the mass density of material must have been far higher and the mean random orbital inclinations and eccentricities must have been much lower.

Subsequent, more detailed accretional simulations (e.g., Davis and Farinella, 1997; Stern and Colwell, 1997a, 1997b; Kenyon and Luu, 1998, 1999; Kenyon and Bromley, 2001) reached similar conclusions and refined an accretion scenario for the formation of the KB that can be summarized as follows. The early KB region between 30 and 50 AU originally contained some 10–50 M_{Earth} of material in the form of dust and small debris. The region was dynamically cold, so that collisions promoted growth. It was likely located somewhat closer to the Sun, perhaps by as much as 5–10 AU, as testified to by Neptune's now well-established migration (e.g., Malhotra, 1993). After of order 100–300 My, objects of size 30–50 km had emerged on nearly circular orbits. These objects were of sufficient size to accelerate their own growth by gravitational runaway. During this mid-stage accretion, objects of size 100–2000 km were rapidly grown. At this point, dynamical excitation of the region was terminated, either by the nearby growth of Neptune, or by Mars-sized objects either grown or scattered into the region. This terminated growth by transforming the accretional environment to an

erosional one, with characteristic collision velocities of order $1\text{--}2 \text{ km s}^{-1}$, as are observed in the KB today. The onset of erosion initiated a fratricide among small bodies, which resulted in the loss of much mass to grinding followed by radiation transport out of the region (Stern and Colwell, 1997b). (Bodies hundreds of kilometers in diameter and larger, by virtue of their mass and gravitational binding energy, were relatively immune to this fratricide.) It was estimated that over 4 Gy, erosional mass loss could deplete the KB down to a few tenths of an Earth mass of material. Although this is higher than the observed KB mass by a factor of a few to perhaps 30 (in the extreme), dynamically driven losses have been shown to be capable of further mass depletion by a factor of 2 to several, thereby making this scenario viable. In the *in situ* accretion scenario (ISAS), plutinos, which migrated to their current orbits after being trapped in Neptune MMRs during their outward migration, preferentially formed at somewhat closer distances (e.g., 20–30 AU) than nonresonant CKBOs and SKBOs.

The salient features of this ISAS are that (1) the early-KB region initially contained an amount of mass consistent with a simple extrapolation of the surface mass density of the giant planet region out to the 20–50 AU zone; and (2) that growth proceeded until excitation by external (or perhaps internal) perturbers interceded, some time between 0.1 and 1 Gy after the formation of the protosolar disk.

An alternative to the ISAS described just above has recently developed (e.g., Morbidelli, 2006). This scenario, called the ‘push out scenario’ (POS), was motivated by dynamicists attempting to re-create various aspects of the orbital distribution of KBOs. In the POS, all of the objects in the KB were accreted in the giant planets region (5–25 AU), rather than beyond this region as the ISAS assumes. These objects were then subsequently transported outward by a combination of gravitational clearing (i.e., scattering) by the giant planets, and resonance migration driven by the outward migration of Neptune to conserve angular momentum during the clearing of the giant planets zone.

The ISAS and POS each have strengths and weaknesses, but a detailed discussion of such is beyond the scope of this chapter. So we restrict our discussion to the point that while the POS points to a compact solar nebula, initially truncated inside $\sim 20 \text{ AU}$, the ISAS points to a broader, dynamically cold disk, stretching out to distances of $\sim 50 \text{ AU}$, or even much farther.

Therefore, a distinguishing test between these scenarios surrounds the issue of whether there is a dynamically cold population of KBOs or smaller debris lying farther from the Sun than the $\sim 50 \text{ AU}$ cutoff of the cold KBO population now observed. Existing observational constraints make clear there is no such cold disk of large KBOs between ~ 50 and $\sim 90 \text{ AU}$. However, this could be the result of a dynamical clearing of this region due to the former presence of a passing star or large planets temporarily scattered there during outer solar system clearing, or due to a change in the accretional environment beyond 50 AU that stunted the growth of bodies there at much smaller sizes than the KBOs in the 30–50 AU zone reached (making such bodies less detectable). Future deep surveys will be required to resolve the question of whether any substantial cold disk lies beyond $\sim 50 \text{ AU}$. Such work must constrain both the population density of small (e.g., 1–10 km) bodies on low-eccentricity, low-inclination orbits in the 50–100 AU zone and the population of large (e.g., $>100 \text{ km}$) bodies on low-eccentricity, low-inclination orbits in the region beyond 100 AU.

10.16.5 Future Observational Goals and Prospects

Existing KBO data and models provide rich insights into this region of the solar system, but, realistically, this field of planetary research is still in its infancy. In the next couple of decades, we expect to see far faster growth in our knowledge and understanding of the KB than in the past, driven by new ground-based observatories, space-based telescopes, space missions (see Chapters 10.17 and 10.18), and computer and detector advances. In this section, we examine some of the highest priority objectives for future KB studies and then discuss the expected advances in ground-based and space-based observations that are expected in the next decade.

10.16.5.1 Goals

The broad goals for the next stage of observational study of KBOs are relatively straightforward. They are to (1) develop a broad understanding of the various populations in the KB and (2) to determine the evolutionary processes that have shaped KBOs to their current state. The first of these goals requires collecting large surveys of albedo, color, composition, shape, and density information. It also requires

thorough searches for KBO atmospheres and the determination of KBO satellite size–frequency–orbital types. It is also strongly desired to begin studying surface heterogeneity on KBOs other than Pluto; this includes albedo, color, and compositional heterogeneity. An additional goal should be to better define the KBO size–frequency histograms down to sizes of kilometers for each orbit type in the KB.

Of particular interest, we point out that binary and higher-order multiple systems offer great potential as tools for studying the physical properties of their components. Knowledge of the orbit of the secondary relative to the primary provides a system mass, and component masses can be determined if absolute astrometry is available to provide motions for each component. Relative sizes can be determined from photometry if one assumes equal reflectance, and absolute sizes inferred if one assumes particular values for the albedos. From sizes and masses, densities can be determined, providing information regarding interior properties and constituents. Of particular interest will be mutual events, like those observed between Pluto and Charon from 1985 to 1990. Such events occur when the line of sight from the Earth intersects the orbital plane of the satellite, so that the parent and satellite(s) occult and transit each other. Photometric and spectroscopic observations during these occultations can provide measurements of absolute sizes, probe atmospheres, and identify surface features and properties on scales far smaller than the resolution limit of the telescopes.

10.16.5.2 Future Ground-Based Observatories

Several large-scale survey telescopes with revolutionary capabilities for KB studies are currently in development, notably the Large Synoptic Survey Telescope (LSST) and the Panoramic Survey Telescope and Rapid Response System (Pan-STARRS). These projects are designed to observe the entire sky observable from the site at a daily or weekly frequency.

The significant new aspect of both these projects compared to previous surveys is to provide regular monitoring of ‘variation’ of objects over the full sky. These projects will provide revolutionary new capabilities for surveying the content of the KB, obtaining large numbers of accurate orbits, and studying KBO light curves, phase curves, colors, and satellite orbits.

The LSST (e.g., Tyson, 2002; Davison and Angel, 2002) is an 8.4 m telescope that will survey the entire available sky every three nights with multiple color filters. Its 3.5 degree CCD field of view will generate 3 billion pixels per image with a plate scale that will be close to 0.2 arcsec/pixel. Some 30 terabytes of data will be produced per night, requiring massive storage and a fast data-processing pipeline. LSST’s performance specifications above indicate it will accomplish a complete survey of KBOs brighter than 25th magnitude every three nights, which will find \sim 50 000 KBOs on the order of 100 km in diameter. Potential observatory sites have been narrowed to three: Cerro Pachon, Las Campanas (both in Chile), and San Pedro Martir (in Mexico). The project goal is to begin operations in 2012.

Pan-STARRS (Kaiser *et al.*, 2002; Jewitt, 2003; Hodapp *et al.*, 2004) will take a different approach to a similar goal. It will consist of four 1.8 m telescopes coordinated to observe the same part of the sky simultaneously with a roughly 3 square degree field of view and a plate scale of 0.3 arcsec. Exposure times of 30–60 s are expected to reach a limiting magnitude of 24. The sky from its site in Hawaii will be observed three times during the darkest nights of each lunation. The first prototype telescope for Pan-STARRS had first light in early 2006, and the full system is planned to be in operation before 2010.

After 3–5 years of operations, it is expected that these projects will produce a complete, flux-limited sample of KBOs that should be free of observational biases. Many tens of thousands of new KBOs should be discovered. This will yield a census of the populations of all known and any major as-yet-undiscovered dynamical groups, as well as substructures therein. They will also provide a vast database of KBO light curves, phase curves, colors, and other data, as noted above.

Smaller-scale ‘pencil beam’ and wide-area surveys to even greater depth using large telescopes will continue to be necessary to probe the extremes of the size distribution of objects in the KB.

Another approach will be taken by TAOS, the Taiwan–America Occultation Survey (Alcock *et al.*, 2003). The goal of TAOS is to sample the low end of the size distribution, by detecting occultations of stars by KBOs with sizes down to a few kilometers. TAOS is designed to compensate for the weakness of doing direct detection of KBOs by reflected light, in which the brightness of the objects decreases approximately by its distance to the fourth power owing to the

compound inverse square law of reflected light sources. TAOS will employ three small (20 inch) robotic telescopes with 2 degree fields of view. They monitor a fixed set of stars throughout the year, observing the same areas of sky simultaneously, and possible events are cross-compared to reject false detections. The third TAOS telescope was installed in April 2004, but the system is still in test as of late 2006. Although detections of asteroid occultations have been made, no definitive KBO occultation events have been detected.

Other ground-based advances to look forward to in the coming 10–15 years include:

- occultation campaigns on known KBOs to better determine their sizes and albedos;
- photometric and spectroscopic campaigns to study mutual event seasons for KBOs with satellites; and
- spectroscopy of KBOs using a new generation of very large optical telescopes coming on line near the middle to end of the 2010–20 decade.

10.16.5.3 Future Space-Based Observations

Two new major space mission capabilities are now under development. We discuss both here. We do not describe missions in study phase or proposals for future missions.

The HST will be able to continue obtaining observations for several more years. If a space shuttle servicing mission occurs as planned in 2008, it will install new gyros, batteries, and a fine guidance sensor, with two new instruments – the Wide-Field Camera 3 (WFC3) and the Cosmic Origins Spectrograph (COS). WFC3 will have greater sensitivity than WFC2, allowing for improved photometry and imaging of fainter objects. COS will restore HST's UV spectroscopy capabilities, and extend them to sources a factor of about 20 times fainter than other HST spectrographs allowed.

The James Webb Space Telescope (JWST), formerly the Next Generation Space Telescope (NGST), will be a large, IR-optimized space telescope to follow HST. JWST is expected to be launched to the second Lagrange point (L2) in 2013–15, and have an expected lifetime of 5–10 years. It will have a 6.5 m diameter mirror (made of 18 hexagonal segments) and an IR camera and spectrograph, providing higher resolution and perhaps a

couple or more orders of magnitude greater sensitivity than HST. Studies by the NGST working group estimate that 10 km diameter objects could be detected at distances of 50 AU. However, JWST will not be able to observe at long enough wavelengths (for the thermal peak around 75 μm) to determine definitive albedos and sizes of KBOs, as can be done with the Spitzer Space Telescope (Emery *et al.*, 2004; Stansberry *et al.*, 2004; Grundy *et al.*, 2005). JWST will be able to study the IR spectra of numerous KBOs and their satellites.

10.16.5.4 Space Missions to the KB and Centaurs

NASA's *New Horizons* mission (Stern and Cheng, 2002) to Pluto and the KB successfully launched on 19 January 2006. After a gravity assist flyby of Jupiter in March 2007, the spacecraft will make its closest approach flyby of the Pluto system on 14 July 2015. *New Horizons* will then go on to conduct one or possibly two KBO encounters in the succeeding 5–6 years. After that time, the spacecraft will emerge from the main region of the KB and into the depleted region beyond 50 AU. KBO targets will be chosen in 2014.

New Horizons carries a suite of seven instruments that include visible light images, IR and UV mapping spectrometers, radio science and radiometry, and *in situ* plasma and dust detectors.

Its primary objectives at each flyby target are to map surface compositions and temperatures; characterize geology and morphology; characterize (or search for) the neutral atmospheres and study escape rates. Secondary objectives include searches for new satellites and rings, ionospheric studies, stereo mapping, and refinement of bulk parameters such as masses, densities, and orbits.

No missions to centaur bodies are in development yet, but *New Horizons* will conduct a distant (\sim 2.8 AU) flyby of Centaur 2002 GO9 in 2009–10, allowing the first-ever measurement of a phase curve for studying the photometric and Bond albedo of such a body.

No additional missions to bodies in the KB are currently in development, but the US Decadal Survey of Planetary Science did place high priority on exploring the diversity of KBOs with space missions. Hence, it would not be surprising to see proposals for future missions.

References

- Alcock C, Dave R, Giamarco J, et al. (2003) TAOS: The Taiwanese–American Occultation Survey. *Earth Moon and Planets* 92: 459–464.
- Altenhoff WJ, Chini R, Hein H, et al. (1988) First radio astronomical estimate of the temperature of Pluto. *Astronomy and Astrophysics* 190: L15–L17.
- Barkume KM, Brown ME, and Schaller EL (2006) Water ice on the satellite of Kuiper Belt object 2003 EL61. *Astrophysical Journal* 640: L87–L89.
- Barucci MA, de Bergh C, Cuby J-G, Le Bras A, Schmitt B, and Romon J (2000) Infrared spectroscopy of the Centaur 8405 Asbolus: First observations at ESO-VLT. *Astronomy and Astrophysics* 357: L53–L56.
- Barucci MA, Doressoundiram A, Tholen D, Fulchignoni M, and Lazzarin M (1999) Spectrophotometric observations of Edgeworth–Kuiper Belt objects. *Icarus* 142: 476–481.
- Barucci MA, Fulchignoni M, Birlan M, Doressoundiram A, Romon J, and Boehnhardt H (2001) Analysis of trans-Neptunian and Centaur colours: Continuous trend or grouping? *Astronomy and Astrophysics* 371: 1150–1154.
- Barucci MA, Lazzarin M, and Tozzi GP (1999) Compositional surface variety among the Centaurs. *Astronomical Journal* 117: 1929–1932.
- Barucci MA, Romon J, Doressoundiram A, and Tholen DJ (2000) Compositional surface diversity in the trans-Neptunian objects. *Astronomical Journal* 120: 496–500.
- Bernstein GM, Trilling DE, Allen RL, Brown ME, Holman M, and Malhotra R (2004) The size distribution of trans-Neptunian bodies. *Astronomical Journal* 128: 1364–1390.
- Bertoldi F, Altenhoff W, Weiss A, Menten KM, and Thum C (2006) The trans-Neptunian object UB313 is larger than Pluto. *Nature* 439: 563.
- Binzel RP (1988) Hemispherical color differences on Pluto and Charon. *Science* 241: 1070–1072.
- Binzel RP (1989) Pluto–Charon mutual events. *Geophysical Research Letters* 16: 1205–1208.
- Binzel RP, Tholen DJ, Tedesco EF, Buratti BJ, and Nelson RM (1985) The detection of eclipses in the Pluto–Charon System. *Science* 228: 1193–1195.
- Boehnhardt H, Tozzi GP, Birkle K, et al. (2001) Visible and near-IR observations of trans-Neptunian Objects. Results from ESO and Calar Alto Telescopes. *Astronomy and Astrophysics* 378: 653–667.
- Brown ME (2001) The inclination distribution of the Kuiper belt. *Astronomical Journal* 121: 2804–2814.
- Brown ME and Calvin WM (1999) Evidence for crystalline water and ammonia ices on Pluto's satellite Charon. *Science* 287: 107–109.
- Brown ME and Koresko CC (1998) Detection of water ice on the centaur 1997 CU 26. *Astrophysical Journal Letters* 505: L65–L67.
- Brown ME, Van Dam M, Bouchez AH, et al. (2006) Satellites of the largest Kuiper Belt objects. *Astrophysical Journal* (in press).
- Brown RH, Cruikshank DP, and Pendleton Y (1999) Water ice on Kuiper Belt object 1996 TO66. *Astrophysical Journal* 519: L101–L104.
- Buie MW, Cruikshank DP, Lebofsky LA, and Tedesco EF (1987) Water frost on Charon. *Nature* 329: 522–523.
- Buie MW and Grundy WM (2000) The distribution and physical state of H₂O on Charon. *Icarus* 148: 324–339.
- Buie MW, Grundy WM, Young EF, Young LA, and Stern SA (2006) Orbits and photometry of Pluto's satellites: Charon, S/2005 P1 and S/2005 P2. *Astronomical Journal* (submitted).
- Buie MW, Tholen DJ, and Wasserman LH (1997) Separate light curves of Pluto and Charon. *Icarus* 125: 233–244.
- Bus SJ, A'Hearn MF, Schleicher DG, and Bowell E (1991) Detection of CN emission from (2060) Chiron. *Science* 251: 774–777.
- Capria MT, Coradini A, De Sanctis MC, and Orosei R (2000) Chiron activity and thermal evolution. *Astronomical Journal* 119: 3112–3118.
- Choi Y, Cohen M, Merk R, and Prialnik D (2002) Long-term evolution of objects in the Kuiper Belt Zone – Effects of insolation and radiogenic heating. *Icarus* 160: 300–312.
- Christy JW and Harrington RS (1978) The satellite of Pluto. *Astronomical Journal* 83: 1005.
- Cohen CJ and Hubbard EC (1965) Libration of the close approaches of Pluto to Neptune. *Astronomical Journal* 70: 10.
- Cook JC, Desch SJ, Roush TL, Geballe TR, and Trujillo CA (2006) Near-infrared spectroscopy of Charon: Possible evidence for cryovolcanism on Kuiper Belt objects. *AAS/DPS Meeting*, 21.02.
- Cruikshank DP, Mason RE, Dalle Ore CM, et al. (2006) Ethane on Pluto and Triton. *AAS/DPS Meeting*, 21.03.
- Cruikshank DP, Pilcher CB, and Morrison D (1976) Pluto – Evidence for methane frost. *Science* 194: 835–837.
- Cruikshank DP, Roush TL, Bartholomew MJ, et al. (1998) The composition of Centaur 5145 Pholus. *Icarus* 135: 389–407.
- Cruikshank DP, Stansberry JA, Emery JP, et al. (2005) The high-albedo Kuiper Belt object (55565) 2002 AW197. *Astrophysical Journal* 624: L53–L56.
- Davies J (2001) *Beyond Pluto Exploring the Outer Limits of the Solar System*. Cambridge: Cambridge University Press.
- Davis DR and Farinella P (1997) Collisional evolution of Edgeworth–Kuiper Belt objects. *Icarus* 125: 50–60.
- Davison WB and Angel JRP (2002) Large synoptic survey telescope mechanical structure and design. In: Anthony TJ and Sidney W (eds.) *Survey and Other Telescope Technologies and Discoveries Proceedings of the SPIE*, 4836, pp. 104–110.
- Delsanti A and Jewitt D (2006) The solar system beyond the planets. In: Blondel Ph and Mason J (eds.) *Solar System Update*. Chichester: Praxis-Springer (in press).
- Delsanti AC, Boehnhardt H, Barrera L, Meech KJ, Sekiguchi T, and Hainaut OR (2001) BVRI photometry of 27 Kuiper Belt objects with ESO/very large telescope. *Astronomy and Astrophysics* 380: 347–358.
- De Sanctis MC, Capria MT, and Coradini A (2001) Thermal evolution and differentiation of Edgeworth–Kuiper Belt objects. *Astronomical Journal* 121: 2792–2799.
- Dobrovolskis AR, Peale SJ, and Harris AW (1997) Dynamics of the Pluto–Charon binary. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, p. 159. Tucson (USA): University of Arizona Press.
- Doressoundiram A, Barucci MA, Tozzi GP, et al. (2005) Spectral characteristics and modeling of the trans-Neptunian object (55565) 2002 AW197 and the Centaurs (55576) 2002 GB10 and (83982) 2002 GO9: ESO Large Program on TNOs and Centaurs. *Planetary and Space Science* 53: 1501–1509.
- Douté S, Schmitt B, Quirico E, et al. (1999) Evidence for methane segregation at the surface of Pluto. *Icarus* 142: 421–444.
- Duffard R, Lazzaro D, Pinto S, et al. (2002) New activity of Charon: Results from 5 years of photometric monitoring. *Icarus* 160: 44–51.
- Duncan M, Quinn T, and Tremaine S (1987) The formation and extent of the solar system comet cloud. *Astronomical Journal* 94: 1330–1338.
- Duncan M, Quinn T, and Tremaine S (1988) The origin of short-period comets. *Astrophysical Journal* 328: L69–L73.
- Durda DD and Stern SA (2000) Collision rates in the present-day Kuiper Belt and Centaur regions: Applications to surface

- activation and modification on comets, Kuiper Belt objects, Centaurs and Pluto–Charon. *Icarus* 145: 220–229.
- Edgeworth KE (1943) The evolution of our planetary system. *Journal of the British Astronomical Association* 53: 181–188.
- Edgeworth KE (1949) The origin and evolution of the solar system. *Monthly Notices of the Royal Astronomical Society* 109: 600–609.
- Elliot JL, Ates A, Babcock BA, et al. (2003) The recent expansion of Pluto's atmosphere. *Nature* 424: 165–168.
- Elliot JL, Dunham EW, Bosh AS, et al. (1988) Stellar occultation observations of Pluto's atmosphere. *Bulletin of the American Astronomical Society* 20: 805.
- Elliot JL, Dunham EW, Bosh AS, et al. (1989) Pluto's atmosphere. *Icarus* 77: 148–170.
- Elliot JL, Olkin CB, Dunham EW, et al. (1995) Jet-like features near the nucleus of Chiron. *Nature* 373: 46.
- Elliot JL, Person MJ, Gulbis AA, Adams ER, et al. (2006) The size of Pluto's atmosphere As revealed by the 2006 June 12 Occultation. *AAS/DPS Meeting*, 21.02.
- Elliot JL and Young LA (1992) Analysis of stellar occultation data for planetary atmospheres: I: Model fitting, with application to Pluto. *Astronomical Journal* 103: 991–1015.
- Emery JP, Cruikshank DP, Van Cleve J, and Stansberry JA (2004) Spitzer/IRS observations of asteroids, Centaurs, and Kuiper Belt objects. *AGU Fall Meeting Abstracts*, A978.
- Eshelman VR (1989) Pluto's atmosphere: Models based on refraction, inversion, and vapor-pressure equilibrium. *Icarus* 80: 439–443.
- Farinella P, Davis DR, and Stern SA (2000) Formation and collisional evolution of the Edgeworth–Kuiper belt. In: Mannings V, Boss AP, and Russell SS (eds.) *Protostars and Planets IV*, pp. 1255–1280. Tucson: University of Arizona Press.
- Farnham TL (2001) The rotation axis of the Centaur 5145 Pholus. *Icarus* 152: 238–245.
- Fernandez JA and Ip WH (1984) Some dynamical aspects of the accretion of Uranus and Neptune – The exchange of orbital angular momentum with planetesimals. *Icarus* 58: 109–120.
- Fernández YR, Jewitt DC, and Sheppard SS (2002) Thermal properties of Centaurs Absolus and Chiron. *Astronomical Journal* 123: 1050–1055.
- Fink and DiSanti (1988) The separate spectra of Pluto and its satellite Charon. *Astronomical Journal* 95: 229–236.
- Fix JD, Neff JS, and Kelsey LA (1970) Spectrophotometry of Pluto. *Astronomical Journal* 75: 895.
- Foster MJ, Green SF, McBride N, and Davies JK (1999) Detection of water ice on 2060 Chiron. *Icarus* 141: 408–410.
- Gerakines PA, Moore MH, and Hudson RL (2001) Energetic processing of laboratory ice analogs: UV photolysis versus ion bombardment. *Journal of Geophysical Research* 106: 33381–33386.
- Gil-Hutton R (2002) Color Diversity among Kuiper Belt objects: The collisional resurfacing model revisited. *Planetary and Space Science* 50: 57.
- Greenberg JM (1982) In: Wilkening LL (ed.) *Comets*, pp. 131–163. Tucson: University of Arizona Press.
- Grundy WM and Buie MW (2001) Distribution and evolution of CH₄, N₂, and CO Ices on Pluto's surface: 1995 to 1998. *Icarus* 153: 248–263.
- Grundy WM and Fink U (1996) Synoptic CCD spectrophotometry of Pluto over the past 15 years. *Icarus* 124: 329–343.
- Grundy WM, Noll KS, and Stephens DC (2005) Diverse albedos of small trans-Neptunian objects. *Icarus* 176: 184–191.
- Grundy WM, Spencer JA, Noll KS, et al. (2006) The orbit, mass, size, albedo, and density of (65489) 2003 FX128: A tidally evolved Centaur. *Icarus* (submitted).
- Grundy WM, Spencer JR, Stansberry JA, et al. (2005) Spitzer/MIPS survey of classical Kuiper Belt objects and a Neptune Trojan. *AAS/Division for Planetary Sciences Meeting Abstracts*, 37.
- Hainaut OR, Delahodde CE, Boehnhardt H, et al. (2000) Physical properties of TNO 1996 TO66. Light curves and possible cometary activity. *Astronomy and Astrophysics* 356: 1076–1088.
- Hainaut OR and Delsanti AC (2002) Colors of minor bodies in the outer solar system. A statistical analysis. *Astronomy and Astrophysics* 389: 641–664.
- Hartmann WK, Cruikshank DP, and Tholen DJ (1989) 2060 Chiron. *Icarus* 83: 1–15.
- Hodapp KW, Kaiser N, Aussel H, et al. (2004) Design of the Pan-STARRS telescopes. *Astronomische Nachrichten* 325: 636–642.
- Hubbard WB, Hunten DM, Dieters SW, Hill KM, and Watson RD (1988) Occultation evidence for an atmosphere on Pluto. *Nature* 336: 452–454.
- Hubbard WB, Yelle RV, and Lunine JI (1990) Non-isothermal Pluto atmosphere models. *Icarus* 84: 1–11.
- Hudson RL and Moore MH (2001) Radiation chemical alterations in solar system ices: An overview. *Journal of Geophysical Research* 106: 33275–33284.
- Hussman H, Sohl F, and Spohn T (2006) Subsurface oceans and deep interiors of medium-sized outer planet satellites and large trans-Neptunian objects. *Icarus* 185: 258–273.
- Jewitt D (2003) Project Pan-STARRS and the outer solar system. *Earth Moon and Planets* 92: 465–476.
- Jewitt D and Luu J (1992) 1992 QB1. *International Astronomical Union Circular* 5611.
- Jewitt D and Luu J (1998) Optical-infrared spectral diversity in the Kuiper Belt. *Astronomical Journal* 115: 1667–1670.
- Jewitt DC (1994) Heat from Pluto. *Astronomical Journal* 107: 372–378.
- Jewitt DC (2002) From Kuiper Belt object to cometary nucleus: The missing ultra-red matter. *Astronomical Journal* 123: 1039–1049.
- Jewitt DC and Luu J (2004) Crystalline water ice on the Kuiper Belt object (50000) Quaoar. *Nature* 432: 731–733.
- Jewitt DC and Luu JX (2001) Colors and spectra of Kuiper Belt objects. *Astronomical Journal* 122: 2099–2114.
- Kaiser N, Aussel H, Burke BE, et al. (2002) Pan-STARRS: A large synoptic survey telescope array. In: Anthony TJ and Sidney W (eds.) *Survey and Other Telescope Technologies and Discoveries*, Proceedings of the SPIE, 4836, pp. 154–164.
- Kenyon SJ and Bromley BC (2001) Gravitational stirring in planetary debris disks. *Astronomical Journal* 121: 538–551.
- Kenyon SJ and Luu JX (1998) Accretion in the early Kuiper Belt: I: Coagulation and velocity evolution. *Astronomical Journal* 115: 2136–2160.
- Kenyon SJ and Luu JX (1999) Accretion in the early Kuiper Belt: II: Fragmentation. *Astronomical Journal* 118: 1101–1119.
- Kowal CT, Dressler A, Adams R, et al. (1977) 1977 UB (slow-moving object kowal). *IAU Circular* 3134.
- Kuiper GP (1951) On the origin of the solar system. In: Hynek JA (ed.) *Proceedings of a Topical Symposium, Commemorating the 50th Anniversary of the Yerkes Observatory and Half a Century of Progress in Astrophysics*, pp. 357. New York: McGraw-Hill.
- Lellouch E, Laureijs R, Schmitt B, et al. (2000) Pluto's Non-isothermal Surface. *Icarus* 147: 220–250.
- Levison HF and Stern SA (2001) On the size dependence of the inclination distribution of the main Kuiper Belt. *Astronomical Journal* 121: 1730–1735.
- Licandro J, Oliva E, and Di Martino M (2001) NICS-TNG infrared spectroscopy of trans-Neptunian objects 2000 EB173 and 2000 WR106. *Astronomy and Astrophysics* 373: L29–L32.

- Licandro J, Pinilla-Alonso N, Pedani M, Oliva E, Tozzi GP, and Grundy WM (2006) The methane ice rich surface of large TNO 2005 FY9: A Pluto-twin in the trans-Neptunian belt? *Astronomy and Astrophysics* 445: L35–L38.
- Luu J and Jewitt D (1996) Color diversity among the Centaurs and Kuiper Belt objects. *Astronomical Journal* 112: 2310.
- Luu JX, Jewitt DC, and Trujillo C (2000) Water ice in 2060 Chiron and its implications for Centaurs and Kuiper Belt objects. *Astrophysical Journal* 531: L151–L154.
- Lykawka PS and Mukai T (2005) Higher albedos and size distribution of large trans-Neptunian objects. *Planetary and Space Science* 53: 1319–1330.
- Malhotra R (1993) The origin of Pluto's peculiar orbit. *Nature* 365: 819–821.
- Marcialis RL, Rieke GH, and Lebofsky LA (1987) The surface composition of Charon – Tentative identification of water ice. *Science* 237: 1349–1351.
- Marcialis RL, Rieke L, and Lebofsky L (1987) The surface composition of Charon: Tentative identification of water ice. *Science* 237: 1349–1351.
- Margot JL, Brown ME, Trujillo CA, and Sari R (2004) HST observations of Kuiper Belt binaries. *AAS/Division for Planetary Sciences Meeting Abstracts* 36.
- McBride N, Green SF, Davies JK, et al. (2003) Visible and infrared photometry of Kuiper Belt objects: Searching for evidence of trends. *Icarus* 161: 501–510.
- McKinnon WB (2002) On the initial thermal evolution of Kuiper Belt objects. In: Barbara Warmbein Proceedings of Asteroids, Comets, Meteors – ACM 2002, ESA SP-500. International Conference, 29 July–2 August 2002, Berlin, Germany, pp. 29–38, ISBN 92-9092-810-7. Noordwijk, The Netherlands: ESA Publications Division.
- McKinnon WB and Barr AC (2006) Structure and evolution of Ice Dwarf Planets. *AAS/DPS Meeting*, 21.01.
- Meech KJ and Belton MJS (1989) (2060) Chiron. *International Astronomical Union Circular* 4770, 1.
- Morbidelli A (2006) Origin and dynamical evolution of comets and their reservoirs. *Astro-ph preprint*.
- Morbidelli A, Levison HF, Tsiganis K, and Gomes R (2005) Chaotic capture of Jupiter's Trojan asteroids in the early Solar System. *Nature* 435: 462–465.
- Noll KS, Stephens DC, Grundy WM, et al. (2002) Detection of two binary trans-Neptunian objects 1997 CQ29 and 2000 CF105, with the Hubble Space Telescope. *Astronomical Journal* 124: 3424–3429.
- Orosei R, Coradini A, de Sanctis MC, and Federico C (2001) Collision-induced thermal evolution of a comet nucleus in the Edgeworth-Kuiper Belt. *Advances in Space Research* 28: 1563–1569.
- Owen TC, Roush TL, Cruikshank DP, et al. (1993) Surface ices and the atmospheric composition of Pluto. *Science* 261: 745–748.
- Pasachoff JM, Souza SP, and Babcock BA, et al. (2005) The structure of Pluto's atmosphere from the 2002 August 21 Stellar Occultation. *Astronomical Journal* 129: 1718–1723.
- Peixinho N, Doressoundiram A, Delsanti A, Boehnhardt H, Barucci MA, and Belskaya I (2003) Reopening the TNOs color controversy: Centaurs bimodality and TNOs unimodality. *Astronomy and Astrophysics* 410: L29–L32.
- Prialnik D, Brosch N, and Ianovici D (1995) Modelling the activity of 2060 Chiron. *Monthly Notices of the Royal Astronomical Society* 276: 1148–1154.
- Rabinowitz DL, Barkume K, Brown ME, et al. (2006) Photometric Observations Constraining the Size, Shape, and Albedo of 2003 EL61, a rapidly rotating, Pluto-sized object in the Kuiper Belt. *Astrophysical Journal* 639: 1238–1251.
- Roe HG (2006) Methane in Pluto's atmosphere. *AAS/DPS Meeting Abstracts* 31.07.
- Sicardy B, Bellucci A, Gendron E, et al. (2006) Charon's size and an upper limit on its atmosphere from a stellar occultation. *Nature* 439: 52–54.
- Sicardy B, Widemann T, Lellouch E, et al. (2003) Large changes in Pluto's atmosphere as revealed by recent stellar occultations. *Nature* 424: 168–170.
- Smith BA, Soderblom LA, Banfield D, et al. (1989) Voyager 2 at Neptune – Imaging science results. *Science* 246: 1422–1449.
- Spencer JR, Stansberry JA, Grundy WM, and Noll KS (2006) ASPIRE IRS survey of hydrated and non-hydrated M asteroids: Preliminary 5–13 micron results. *AAS/DPS Meeting*, 34.01.
- Spencer JR, Stansberry JA, Trafton LM, Young EF, Binzel RP, and Croft SK (1997) Volatile transport, seasonal cycles, and atmospheric dynamics on Pluto. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 435. Tucson: University of Arizona Press.
- Stansberry JA, Cruikshank DP, Grundy WM, et al. (2004) Far-IR photometry of Centaurs and Kuiper Belt objects with Spitzer. *AAS/ Division for Planetary Sciences Meeting Abstracts*, 36.
- Stansberry JA, Grundy WM, Margot JL, et al. (2006) The Albedo, size, and density of binary Kuiper Belt object (47171) 1999 TC36. *Astrophysical Journal* 643: 556–566.
- Stansberry JA, Lunine JI, Hubbard WB, Yelle RV, and Hunten DM (1994) Mirages and the nature of Pluto's atmosphere. *Icarus* 111: 503–513.
- Stephens DC and Noll KS (2005) The high fraction of binaries in the cold classical Kuiper Belt. *DPS Meeting #37*, 56. 16. *American Astronomical Society*.
- Stern SA (1989a) ISM-induced erosion and dynamical drag in the Oort cloud. *Bulletin of the American Astronomical Society* 21: 941.
- Stern SA (1989b) Pluto—Comments on crustal composition, evidence for global differentiation. *Icarus* 81: 14–23.
- Stern SA (1991) On the number of planets in the outer solar system – Evidence of a substantial population of 1000-km bodies. *Icarus* 90: 271–281.
- Stern SA (1992) The Pluto–Charon system. *Annual Review of Astronomy and Astrophysics* 30: 185–233.
- Stern SA (1995) Collisional time scales in the Kuiper disk and their implications. *Astronomical Journal* 110: 856–868.
- Stern SA (1996) On the collisional environment, accretion time scales, and architecture of the massive, primordial Kuiper Belt. *Astronomical Journal* 112: 1203–1211.
- Stern SA (2002) Evidence for a collisional mechanism affecting Kuiper Belt object colors. *Astronomical Journal* 124: 2297–2299.
- Stern SA (2003) The evolution of comets in the Oort cloud and Kuiper Belt. *Nature* 424: 639–642.
- Stern SA (2005) Regarding the accretion of 2003 VB₁₂ (Sedna) and like bodies in distant heliocentric orbits. *The Astronomical Journal* 129: 526–530.
- Stern SA (in press) New horizons: NASA's Pluto–Kuiper Belt missions. *Space Science Reviews*.
- Stern SA, Buie MW, and Trafton LM (1997) HST High-resolution images and maps of Pluto. *Astronomical Journal* 113: 827.
- Stern SA and Cheng A (2002) NASA Plans Pluto–Kuiper Belt mission. *EOS* 83: 108–109.
- Stern SA and Colwell JE (1997a) Accretion in the Edgeworth–Kuiper belt: Forming 100–1000 km radius bodies at 30 AU and beyond. *Astronomical Journal* 114: 841–849.
- Stern SA and Colwell JE (1997b) Collisional erosion in the primordial Edgeworth–Kuiper Belt and the generation of the 30–50 AU Kuiper gap. *Astrophysical Journal* 490: 879–882.
- Stern SA and Trafton L (1984) Constraints on bulk composition, seasonal variation, and global dynamics of Pluto's atmosphere. *Icarus* 57: 231–240.

- Stern SA and Trafton L (in press) On the atmosphere of objects in the Kuiper Belt. In: Barrucci A, Morbidelli A, and Cruikshank D (eds.) *Trans-Neptunian Objects*. LPI Press.
- Stern SA, Trafton LM, and Gladstone GR (1988) Why is Pluto bright? Implications of the albedo and lightcurve behavior of Pluto. *Icarus* 75: 485–498.
- Stern SA, Weaver HA, Steffl AJ, et al. (2006) A giant impact origin for Pluto's small moons and satellite multiplicity in the Kuiper belt. *Nature* 439: 946–948.
- Stern SA, Weintraub DA, and Festou MC (1993) Evidence for a low surface temperature on Pluto from millimeter-wave thermal emission measurements. *Science* 261: 1713.
- Strazzulla G and Johnson RE (1991) Comets in the Post Halley era. In: Newburn RL Jr. (ed.) *Irradiation Effects on Comets and Cometary Debris*, vol. 1, pp. 243–275. The Netherlands: Kluwer Academic Publisher.
- Strobel DF, Zhu X, Summers ME, and Stevens MH (1996) On the vertical thermal structure of Pluto's atmosphere. *Icarus* 120: 266–289.
- Sykes MV, Cutri RM, Lebofsky L A, and Binzel RP (1987) IRAS serendipitous survey observations of Pluto and Charon. *Science* 237: 1336–1340.
- Tegler SC and Romanishin W (1998) Two distinct populations of Kuiper Belt objects. *Nature* 392: 49–51.
- Tegler SC and Romanishin W (2000) Extremely red Kuiper-Belt objects in near-circular orbits beyond 40 AU. *Nature* 407: 979–981.
- Tegler SC and Romanishin W (2003) Resolution of the Kuiper Belt object color controversy: Two distinct color populations. *Icarus* 161: 181–191.
- Thébault P and Doressoundiram A (2003) Colors and collision rates within the Kuiper Belt: Problems with the collisional resurfacing scenario. *Icarus* 162: 27–37.
- Tholen DJ and Buie MW (1997) Bulk Properties of Pluto and Charon. Stern SA and Tholen DJ, (eds.) *Pluto and Charon*, 193pp. Tucson: University of Arizona Press.
- Tholen DJ, Buie MW, Storrs AD, and Lark N (1986) Improved physical parameters for the Pluto–Charon system. *Bulletin of the American Astronomical Society* 18: 821.
- Thomas N, Eggers S, Ip W-H, et al. (2000) Observations of the Trans-Neptunian Objects 1993 SC and 1996 TL66 with the infrared space observatory. *Astrophysical Journal* 534: 446–455.
- Thompson RW, Murray BG JPT, Khare BN, and Sagan C (1987) Coloration and darkening of methane clathrate and other ices by charged particle irradiation: Applications to the outer solar system. *Journal of Geophysical Research* 92: 14933–14947.
- Throop HB (2000) *Light scattering and Evolution of Protoplanetary Disks and Planetary Rings*. PhD Thesis, University of Colorado at Boulder.
- Trafton L (1990) A two-component volatile atmosphere for Pluto. I: The bulk hydrodynamic escape regime. *Astrophysical Journal* 329: 512–523.
- Trafton L and Stern SA (1983) On the global distribution of Pluto's atmosphere. *Astrophysical Journal* 267: 872–881.
- Trafton LM, Hunten DM, Zahnle KJ, and McNutt RL Jr. (1997) Escape processes at Pluto and Charon. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, 475pp. Tucson: University of Arizona Press.
- Trilling DE and Bernstein GM (2006) Light Curves of 20–100 km Kuiper Belt objects using the Hubble Space Telescope. *Astronomical Journal* 131: 1149–1162.
- Trujillo CA and Brown ME (2002) A Correlation between inclination and color in the classical Kuiper Belt. *Astrophysical Journal* 566: L125–L128.
- Trujillo CA, Brown ME, Barkume KM, Schaller EL, and Rabinowitz DL (2006) The Surface of 2003EL61 in the near infrared. *Astrophysical Journal* 655: 1172–1178.
- Tryka KA, Brown RH, Chruikshank DP, Owen TC, Geballe TR, and Debergh C (1994) Temperature of nitrogen ice on Pluto and its implications for flux measurements. *Icarus* 112: 513–527.
- Tyson JA (2002) Large synoptic survey telescope: Overview. In: Anthony TJ and Sidney W (eds.) *Survey and Other Telescope Technologies and Discoveries. Proceedings of the SPIE*, 4836, pp. 10–20.
- Veillet C, Parker JW, Griffin I, et al. (2002) The binary Kuiper-Belt object 1998 WW31. *Nature* 416: 711–713.
- Walker AR (1980) An occultation by Charon. *Monthly Notices of the Royal Astronomical Society* 192: 47P–50P.
- Walker MF and Hardie R (1955) A Photometric determination of the rotational period of Pluto. *Publications of the Astronomical Society of the Pacific* 67: 224.
- Ward WR (1996) Planetary accretion. Completing the inventory of the solar system. *Astronomical Society of the Pacific Conference Proceedings* 107: 337–361.
- Weaver HA, Stern SA, Mutchler MJ, et al. (2006) Discovery of two new satellites of Pluto. *Nature* 439: 943–945.
- Womack M and Stern SA (1999) Detection of carbon monoxide in (2060) Chiron. *Solar System Research (Astronomicheskii Vestnik)* 33: 187.
- Yelle RV and Elliot JL (1997) Atmospheric structure and composition. In: Stern SA and Tholen DJ (eds.) *Pluto and Charon*, pp. 347–390. Tucson: University of Arizona Press.
- Yelle RV and Lunine JI (1989) Evidence for a molecule heavier than methane in the atmosphere of Pluto. *Nature* 339: 288–290.
- Young EF, Binzel RP, and Crane K (2001) A two-color map of Pluto's sub-Charon hemisphere. *Astronomical Journal* 121: 552–561.
- Young EF, Buie MW, French RG, et al. (2006) The detailed vertical structure of Pluto's atmosphere from the 12 Jun 2006 Stellar occultation. *AAS/DPS Meeting*, 31.04.
- Young LA, Elliot JL, Tokunaga A, de Bergh C, and Owen T (1997) Detection of gaseous methane on Pluto. *Icarus* 127: 258–270.

Relevant Websites

- <http://taos.asiaa.sinica.edu.tw> – Census of Kuiper Comets: The Taiwan–America Occultation Survey.
- <http://www.lsst.org> – Large Synoptic Survey Telescope.
- <http://pluto.jhuapl.edu> – *New Horizons* Website.
- <http://pan-starrs.ifa.hawaii.edu/public/> – Pan-STARRS: Panoramic Survey Telescope & Rapid Response System.
- <http://www.boulder.swri.edu> – The *Distant EKOs* Newsletter, Department of Space Studies of the Southwest Research Institute.

10.17 Mission Analysis Issues for Planetary Exploration Missions

Y. Langevin, Institut d' Astrophysique Spatiale, CNRS/Univ. Paris Sud, France

© 2007 Elsevier B.V. All rights reserved.

10.17.1	Introduction	565
10.17.2	Scientific Rationale and Present Status of Planetary Exploration Missions	565
10.17.3	Energy Requirements and Mass Budgets for Planetary Missions	567
10.17.3.1	Leaving the Gravitational Influence of the Earth	567
10.17.3.2	Flyby Missions	570
10.17.3.3	Rendezvous Missions	571
10.17.3.3.1	Rendezvous missions to natural satellites of giant planets	572
10.17.3.3.2	Rendezvous missions to comets and Earth-crossing asteroids	572
10.17.4	Remote-Sensing and <i>In Situ</i> Missions to Venus and Mars	573
10.17.4.1	Orbiter Missions Not Implementing Aerocapture	575
10.17.4.2	<i>In Situ</i> Missions and Orbiter Missions Implementing Aerocapture	576
10.17.5	Orbit Evolution around Planetary Bodies	576
10.17.6	<i>In Situ</i> Missions to Atmosphereless Bodies	578
10.17.7	Gravity-Assist Missions: Giant Planets, Mercury, Asteroid, and Comet Rendezvous	580
10.17.8	Advanced Propulsion Systems: Solar Sail, Ion Propulsion	586
10.17.8.1	Solar Sail Propulsion	587
10.17.8.2	Ion Propulsion	587
10.17.9	The Specific Challenges of Sample Return Missions	589
References		592

10.17.1 Introduction

The chapter presents the scientific rationale and present status of planetary exploration missions in Section 10.17.2. In the third section, the energy requirements and mass budgets of planetary missions are evaluated, first for leaving the Earth, then to visit (flyby missions) or go into orbit (rendezvous missions) around bodies of interest in the solar system. Section 10.17.4 is dedicated to remote-sensing and *in situ* missions to Venus and Mars. Section 10.17.5 discusses the evolution of orbits around planetary bodies. Section 10.17.6 presents the challenges of *in situ* missions to atmosphereless bodies. Section 10.17.7 presents the gravity-assist techniques which are required for high-energy targets (giant planets, Mercury, comet rendezvous). Section 10.17.8 discusses advanced propulsion systems, the solar sail and ion propulsion. In the last section (Section 10.17.9), the specific challenges of sample return missions are discussed.

10.17.2 Scientific Rationale and Present Status of Planetary Exploration Missions

For centuries, major advances in planetary science had to rely on ground-based observations. The increasing size of telescopes up to the present 8-m-class observatories (Keck, VLT), and the increasing sophistication of focal plane instruments have made possible major advances in terms of wavelength coverage, and spectral and spatial resolution for observing solar system bodies. Two recent developments have enhanced much the potential of ground-based observations: the development of adaptive optics (which brings telescopes close to their diffraction limits by correcting for atmospheric turbulence) and interferometry (which increases the effective aperture, improving spatial resolution). New results can therefore be expected from the next generation of ground-based observatories in the 2010–20 time frame, in particular The Atacama Large Millimeter Array (ALMA) (in the

sub-millimeter wavelength range), The Square Kilometer Array (SKA) (at radio wavelengths), and extremely large telescopes (aperture >25 m). Space observatories are not limited to the transmission windows of the Earth's atmosphere. The most prominent examples for planetary science have been the Hubble Space Telescope, Infrared Space Observatory (ISO) and Space Infrared Telescope Facility (SIRTF) in the infrared (IR) range, and Far Ultraviolet Spectroscopic Explorer (FUSE) in the Ultraviolet (UV) range. The next step will be the launch of the Herschel Observatory (sub-millimeter wavelength range) in 2008. An important perspective for planetary science which goes beyond our solar system is the detection and characterization of other planetary systems. A range of missions is already programmed, starting with Convection Rotation and Planetary Transits (COROT) (2006, CNES-ESA) and Kepler (2008, NASA) for the observation of planetary transits, then GAIA (2012, ESA) which should identify thousands of giant planets through the displacement of the central star, and ultimately large interferometers in space which will be able to directly image extra-solar planets as well as characterizing possible atmospheres. Such ambitious programs will most likely require a worldwide collaboration in the 2015–20 time frame.

While recognizing the importance of these contributions, it is clear that most breakthroughs in solar system science have resulted and will continue in the future to result from space missions to other solar system bodies, starting from the landmark Luna III mission to the Moon which revealed for the first time the farside. Space observatories (in orbit around the Earth or at the L2 Lagrangian point of the Earth relative to the Sun) are not limited to the transmission windows of the atmosphere of the Earth. This also applies to remote-sensing planetary missions, which can achieve much higher spatial resolutions as they get much closer to the target body, even with relatively small optical devices (a typical planetary mission payload ranges from 50 to 150 kg in mass). For a discussion of the instrumentation of planetary exploration missions see Chapter 10.18. Let us consider a 10-m-class telescope operating at $1\text{ }\mu\text{m}$. Its maximum resolution is 10^{-7} rad assuming that the diffraction limit is reached using adaptive optics. The best observing conditions are at quadrature for Mercury and Venus (maximum angular distance from the Sun) and at opposition for outer bodies. Distances therefore range from 55 million km (Mars at the best oppositions) to several hundred million km. With a space mission, one can easily

implement a camera with an aperture of at least 10 cm (a 70-cm-aperture camera High-Resolution Imaging Science Experiment (HiRISE) is orbiting Mars since Nov 2006), corresponding to a diffraction-limited resolution of 10^{-5} rad or better. The closest distance for a flyby or an orbiter is typically 300 km for most planets and asteroids, 30 000 km for Jupiter, and 80 000 km for Saturn (outer edge of the A ring). For Mars, the resolution of a camera with a 10–20 kg mass is in the meters range instead of at best 5.5 km from the Earth. Therefore, flyby missions (over a short time interval) and orbiter missions provide improvements by several orders of magnitudes over the best ground-based or Earth-orbiting facilities. Orbiter missions make it possible to monitor temporal variations at high spatial resolution. Such missions also provide unique opportunities for characterizing the environment of planetary bodies in terms of particles, fields, and their interactions with the solar wind. By implementing *in situ* missions (landers, rovers, airborne modules if there is an atmosphere), it is possible to bring to bear the full range of techniques used in Earth sciences: chemical and mineralogical analysis, seismology (which is best implemented with a network of stations), atmospheric sensors, etc. Eventually, sample return missions make it possible to implement powerful analysis techniques which are still required today for very high-precision measurements of trace elements and isotopic compositions (radiochronology, isotopic anomalies). This may change in the future considering the remarkable improvements of *in situ* analysis techniques over the last 20 years (e.g., mass spectrometry with a resolution of 1000 or more, implemented on the Rosetta comet rendezvous mission of European Space Agency (ESA) only 20 years after it became available in the laboratory).

The sequence of missions to a given planetary body in terms of increasing complexity is therefore the following:

1. first exploration by a flyby mission,
2. comprehensive characterization by remote-sensing techniques on orbiter missions,
3. *in situ* investigations (landers, rovers, descent probes, balloons, airplanes, etc.), and
4. sample return missions.

The edition of this book corresponds to a timely period for assessing the status of planetary exploration and outlining the next perspectives. All classes of solar system bodies with the exception of Kuiper Belt objects (KBOs) (beyond Neptune) have been visited

at least once by flyby missions. The last gap will be filled by the New Horizons NASA mission, which will encounter Pluto, one of the largest KBOs, in 2015. Similarly, the remote-sensing phase is well underway, with the highly successful missions to Mars (Mars Global Surveyor MGS), Mars Odyssey, Mars Reconnaissance Orbiter (MRO) for NASA; Mars Express for ESA; the Cassini mission to the Saturn–Titan system, the Venus Express mission to Venus (ESA), the MESSENGER mission (NASA, launched in 2005), and the BepiColombo mission (ESA, launch scheduled in 2013) to Mercury. For small bodies, one can mention the rendezvous missions towards comet Churyumov–Gerasimenko (Rosetta, ESA), near-Earth asteroids Eros (NEAR, NASA), and Itokawa (Hayabusa, Japan) and the main-belt asteroids Vesta and Ceres (DAWN, NASA). *In situ* investigations recently suffered some setbacks (failure of the Beagle 2 Mars lander, descoping of a Mercury lander on BepiColombo) but remarkable successes have been obtained by the Mars rover missions of NASA and the Huygens descent probe in the atmosphere of Titan (ESA), which provided information all the way down to the surface. The Mars Surface Laboratory (MSL) of NASA (to be launched in 2009) will build on the remarkable success of Mars vehicles and orbiter missions, and another ambitious rover mission to Mars can now be expected in the 2011–13 time frame in the framework of the exploration program of ESA. For 30 years, lunar dust and rocks from Apollo and Luna had remained the only samples brought back from another planetary body. Samples of cometary dust have been successfully recovered in early 2006 from comet Wild-1 (Stardust, NASA), with intriguing results on early mixing in the proto-solar nebula. However, it now looks likely that the Japanese Hayabusa mission will not succeed in bringing back a sample of asteroid material. While orbiter missions dedicated to specific goals will continue to be useful in terms of time coverage and relay capability, it is likely that solar system science missions beyond the 2015 time frame will focus on *in situ* investigations and sample returns. These considerations are important for the purpose of this chapter, as the resources required are increasingly larger for each step in planetary exploration, even if the presence of an atmosphere around the target body can in part alleviate for these higher energy requirements. The following sections present the major constraints on mission design for planetary exploration. The basic issues of astrodynamics are presented in Bate

et al. (1971). More recent textbooks have been published by Battin (1999) and Vallado (2001). A discussion of these questions focused on mission design techniques is also available (Langevin, 2005). A review of the most recent advances in mission analysis is provided by Marsden and Ross (2006).

10.17.3 Energy Requirements and Mass Budgets for Planetary Missions

10.17.3.1 Leaving the Gravitational Influence of the Earth

Apart from missions to the Moon, all planetary missions have to leave the gravitational influence of the Earth. In order to compare the different types of missions, one can consider as a starting point a near circular orbit around the Earth at an altitude of ~ 330 km, which lies beyond the upper reaches of the Earth's atmosphere. A large velocity increment can then be provided in as short a time as possible by an upper stage, so as to exceed the escape velocity. This strategy was used for both Mars Express (Jun 2003) and Venus Express (Nov 2005) with the Soyuz launcher and its Fregat upper stage.

At this altitude, the circular velocity V_c is 7.71 km s^{-1} and the escape velocity V_{esc} , which is always $\sqrt{2}$ times larger than V_c , is 10.90 km s^{-1} . Velocity increments lower than 3.19 km s^{-1} provided by the upper stage will therefore result in bound orbits, and velocity increments larger than 3.19 km s^{-1} will result in interplanetary trajectories. The interplanetary departure velocity V_{inf} can easily be derived from energy conservation arguments:

$$V_{\text{inf}}^2 = V_{\text{per}}^2 - V_{\text{esc}}^2 \quad [1]$$

where V_{per} is the velocity at pericenter after the upper-stage burn.

The resources required for each class of trajectory can then be compared from the velocity increment (ΔV) which is required in addition to that which would be needed to reach a bound orbit with a semimajor axis of $320\,000$ km, which goes much beyond that of the Moon (**Table 1**).

It is important to note that the ΔV required from the upper stage to reach the Moon (3.14 km s^{-1}) and to leave the Earth with a 3 km s^{-1} relative velocity (3.61 km s^{-1}) are similar (460 m s^{-1} difference). Very large departure velocities (9 km s^{-1}) require additional velocity increments similar to the 3.19 km s^{-1}

Table 1 ΔV requirements on the upper-stage burn for each class of trajectories

Type of trajectory	V_{per} (km s $^{-1}$)	Upper stage ΔV (km s $^{-1}$)	ΔV over bound orbit (km s $^{-1}$)
Moon-crossing orbit	10.85	3.14	0.00
Marginal escape	10.90	3.19	0.05
Departure at 3 km s $^{-1}$	11.31	3.61	0.46
Departure at 5 km s $^{-1}$	12.00	4.29	1.15
Departure at 7 km s $^{-1}$	12.96	5.25	2.11
Departure at 9 km s $^{-1}$	14.14	6.43	3.29

which are already needed to escape from a circular orbit.

These considerations have a major impact on the mass available for interplanetary modules. This directly derives from the basic physical constraints on rockets, first discussed by Tsiolkovsky in 1903. An upper stage, similarly to any rocket, ejects exhaust gas at a large velocity V_{ej} (~ 3.25 km s $^{-1}$ for the Fregat upper stage, ~ 4.1 km s $^{-1}$ for the Centaur cryogenic upper stage). From momentum balance, we can write

$$M(t) \, dv = dM \, V_{\text{ej}} \quad [2]$$

where $M(t)$ is the mass at a given time during the thrust. Integrating eqn [2] over the full duration of the thrust results in the basic constraint of spaceflight, the ‘rocket equation’ (Tsiolkovsky, 1903)

$$\Delta V = V_f - V_i = V_{\text{ej}} \ln(M_i/M_f) \quad [3]$$

where M_i and M_f are the initial and final mass of the upper stage with its payload, and ΔV is the total velocity increment.

In the case of the Soyuz–Fregat launch strategy used for Mars Express and Venus Express, the equivalent mass which could be injected in a near circular orbit is 7200 kg while the dry mass of the Fregat stage with the payload attachment device is 1150 kg. It should be noted that this ‘useless’ dry mass ends up with the same velocity as the spacecraft itself; hence, it has to be subtracted from M_f so as to obtain the ‘useful’ mass injected into a given orbit.

Figure 1 provides the useful mass as derived from the rocket equation which can be launched at a given departure velocity from the Earth–Moon system with Soyuz–Fregat in its Mars Express configuration. It is clear that there is very little available mass for escape velocities larger than 6 km s $^{-1}$, while departure velocities lower than 3.5 km s $^{-1}$ result in significant mass budgets even with a relatively small (and cheap !) launcher. This is a slightly optimistic view, as it considers that the Fregat burn is performed

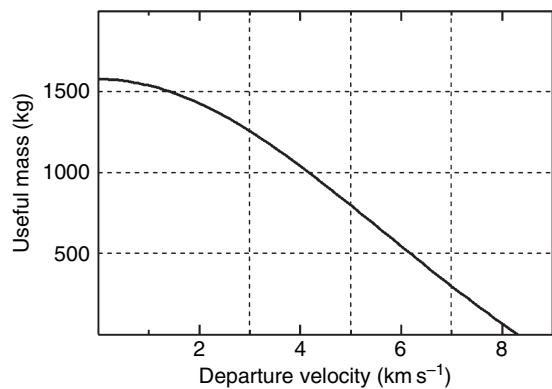


Figure 1 Useful mass as a function of departure velocity with the Soyuz–Fregat combination used for the launch of Mars Express in 2003, not taking into account gravity losses. The departure velocity of Mars Express was ~ 3 km s $^{-1}$; hence, the available useful mass was ~ 1250 kg.

at a 330 km altitude. However, even with the powerful thruster of Fregat (20 000 N, providing an acceleration of 2.8 m s $^{-2}$ at full load, 10 m s $^{-2}$ at the end of the thrust arc), the thrust extends over several hundred seconds so as to reach escape velocity. Part of the thrust is therefore provided when the upper stage is at higher altitudes. Close to escape, the energy provided by each m s $^{-1}$ of ΔV is proportional to the local escape velocity. At an altitude of 2000 km, the local escape velocity is 9.75 km s $^{-1}$ instead of 10.90 km s $^{-1}$ at an altitude of 330 km s $^{-1}$; hence, the efficiency of the thrust in terms of energy has been reduced by 10%. This penalty incurred with extended thrust arcs is called a ‘gravity loss’. It is increasingly severe for very large velocity increments which require longer thrust arcs. As an example, the actual performance of the Soyuz–Fregat drops to zero slightly below 8 km s $^{-1}$, instead of 8.15 km s $^{-1}$ as in Figure 1. In practical terms, a single stage such as Fregat can provide at best a velocity increment which is ~ 1.5 times larger than the exhaust velocity.

For spacecraft propulsion systems, the practical limit is lower, in the range of the exhaust velocity, as the initial mass is that at departure and some useful mass must remain for the scientific mission.

Two alternative launch strategies with an intermediate bound orbit can be considered:

1. For low departure velocities, it is possible to effectively use the Moon, which is a very large satellite when compared to its primary. A Moon gravity-assist (MGA) (see Section 10.17.7) from a bound orbit can provide a departure velocity close to 1.4 km s^{-1} . If the Moon is in a favorable location, this strategy provides for free the launch capability of the Moon-crossing orbit (1610 kg) instead of 1500 kg as in **Figure 1**. There are severe limitations to this strategy, as the departure direction must be close to the orbit plane of the Moon. The 7% improvement in the mass budget is therefore only of interest in very specific cases (ion propulsion missions, see Section 10.17.8).

2. For large departure velocities, it is not effective to carry along the large dry mass of the Fregat upper stage (1150 kg) which then exceeds the useful mass. It is interesting to consider launching as much mass as possible (1610 kg) into a Moon-crossing elliptical orbit. The spacecraft thruster (or a dedicated propulsion stage) can then be used as a new stage at the next perigee. This indirect launch strategy is discussed by Farquhar and Dunham (1999). The required ΔV at perigee can be derived from the fourth column in **Table 1**, which indicates the difference between the perigee velocity of the Moon-crossing orbit and that required for departure at a given velocity. Interplanetary spacecraft thrusters typically provide 400 N, much less than the 20 000 N of Fregat. The exhaust velocity is only slightly smaller (3.1 km s^{-1} instead of 3.28 km s^{-1}). For large departure velocities, the thrust arc extends far before and after perigee, and gravity losses begin to offset the advantage of jettisoning the large Fregat upper stage. For Mars Express (departure at 3 km s^{-1}), the spacecraft thruster has to provide 460 m s^{-1} (actually 500 m s^{-1} due to gravity losses). The final mass after the spacecraft thruster burn is then 1370 kg, 130 kg more than with a direct injection with Fregat. One should however take into account that 240 kg of additional fuel have to be carried, with a spacecraft mass at separation of 1610 kg. This requires larger tanks as well as a more robust structure than the ‘direct injection’ spacecraft with its 1240 kg. For each additional kilogram of fuel, 0.1–0.14 kg of tanks and structure have to be added to

the spacecraft (this ratio is called the ‘tankage factor’). With 240 kg of additional fuel, the penalty in dry mass is ~ 30 kg, hence the benefit of such a launch strategy is reduced to 100 kg, which is still quite significant. For a departure velocity of 6 km s^{-1} , the bound-orbit launch strategy increases the available mass after departure from 500 to 700 kg, even taking into account gravity losses. For departure velocities ranging from 3 to 6 km s^{-1} , there is therefore a complex tradeoff between mass budget, complexity, cost, and risk: the ‘direct injection’ spacecraft is lighter, hence cheaper, and a major maneuver 1 month after launch may not be desirable. A bound-orbit launch extends the mission duration by 1–2 months, but extending the launch window and coping with launcher errors becomes very cheap. Finally, it is important to note that there is a practical limit of $\sim 3 \text{ km s}^{-1}$ on onboard ΔV , as this corresponds to the ejection velocity, hence more than 70% of the spacecraft mass would consist of fuel, tanks, and the corresponding structural mass. Therefore, the bound orbit launch strategy cannot be considered for missions requiring large ΔV capabilities after departure.

Soyuz–Fregat is expected to be launched from Kourou starting in 2009, and new versions of Fregat are being developed. A Moon-crossing orbit capability of 2140 kg with Soyuz–Fregat is therefore considered as the baseline for the BepiColombo mission to Mercury (to be launched in 2013). For missions requiring more mass, one has to use a more capable launcher. Ariane 5 (ECA version), Atlas 5 (421), Delta 4 (M+), or Proton provides a Moon-crossing orbit capability larger than 4500 kg (3500 kg or more for a departure at 3 km s^{-1}). With such launchers, a direct injection is therefore likely to meet the mass requirements of most solar system exploration missions. The departure velocity is however still limited to $\sim 7 \text{ km s}^{-1}$ if the upper stage is not cryogenic (Ariane 5, Proton), as the same equations as those for the Fregat upper stage of Soyuz apply. NASA has a major asset for high-energy missions as it was able to develop a cryogenic upper stage (Centaur) as early as 1963. The ejection velocity of a H_2/O_2 engine is larger than 4 km s^{-1} , compared to 3.1 km s^{-1} for bipropellant thrusters such as that of Fregat, and the decrease in useful mass with departure velocity is slower. This made it possible to implement direct missions to Jupiter, which require departure velocities of at least 9 km s^{-1} : Pioneer 10/11, launched by Atlas/Centaur in

1972/1973, and Voyager 1/2, launched by Titan/Centaur in 1977. Such direct missions to Jupiter have a very short cruise time (1.75 years to Jupiter). However, the useful mass is low: the mass of Pioneer 10 was only 270 kg, hence this strategy has until now only been used for flyby missions to the outer solar system, the most recent being ‘New Horizons’ (NASA), launched in 2006 towards Pluto and the Kuiper Belt.

10.17.3.2 Flyby Missions

Flyby missions only require that the spacecraft leaving Earth crosses the path of the target body. The Moon is a specific case, as it orbits the Earth. It can be visited by a spacecraft on a bound geocentric orbit, with a velocity at perigee lower than the escape velocity (see [Table 1](#)). Near-Earth asteroids and several comets have orbits which come quite close to that of the Earth. The velocities relative to the Earth which are required to reach these bodies (albeit at a high flyby velocity) can be so low ($<1.5 \text{ km s}^{-1}$) that a simple lunar swingby from a bound orbit is adequate. The mass budget is then similar to that of a lunar flyby mission. The best example of such a cheap flyby mission is the magnetospheric mission ISEE-3 of NASA, retargeted and renamed International Comet Explorer (ICE) after a lunar flyby which made it possible to encounter comet Giacobini-Zinner in 1991 ([Farquhar, 2001](#)).

Multiple flybys of comets and asteroids can be cheaply implemented using gravity-assist strategies (see [Section 10.17.7](#)), thereby enhancing the scientific return ([Farquhar et al., 2002](#)). This usually requires retargeting the spacecraft by flying close to a planet. When this planet is the Earth, the period of each successive orbit has to be close to a simple fraction of 1 year. As an example, the trajectory of the ESA probe GIOTTO on its way to the flyby of comet Halley in 1986 had a orbital period close to 0.75 years, which brought the spacecraft back to the Earth after four orbits and 3 years. It was then possible to adjust the Earth swingby so as to visit comet Schwassmann-Wachman 3 in 1991. Similarly, USSR’s VEGA-1 and VEGA-2 missions used a swingby of Venus (with the deployment of balloons in the atmosphere) to reach comet Halley at nearly the same time as GIOTTO. The most comprehensive mission of this type was expected to be NASA’s CONTOUR. Four flybys of comets and asteroids were planned in the nominal mission, three intermediate Earth swingbys retargeting the spacecraft towards the next target. This mission unfortunately failed at launch in 2001.

All other planetary bodies of interest (planets, natural satellites apart from the Moon, main-belt asteroids, most comets) do not come close to the orbit of the Earth. A transfer trajectory is then required, with a significant departure velocity. Apart from near-Earth objects, minor bodies consist of three main groups: comets, main-belt asteroids, and KBOs which orbit beyond Neptune. Pluto is not a planet, but one of the largest representatives of KBO’s. Most minor bodies have large eccentricities and large inclinations. Such characteristics much increase the cost of a flyby mission, and even more so that of a rendezvous mission. Some large main-belt asteroids, such as 4 Vesta, have relatively low eccentricity and inclination, which is also the case for the seven major planets. Visiting a large natural satellite with a flyby mission has very similar requirements to that of a mission to the planet itself.

A first approximation of the relative complexity of solar system exploration missions can be made by assuming that all planetary orbits are circular, with an orbit in the same plane as the orbit of the Earth. The problem in this coplanar, circular approximation becomes very simple: the cheapest transfer between the Earth and such an idealized planet is an ellipse which is tangent to the two circles (Hohmann transfer). This applies both to a flyby mission and a rendezvous mission. All velocities in the solar system can be scaled to the orbital velocity of the Earth, $V_e = 29.78 \text{ km s}^{-1}$, if the semimajor axis of the Earth is considered as the unit of distance (astronomical unit, or AU). If R_p is the semimajor axis of the planet in these units, we have

$$A = (1 + R_p)/2 \quad [4]$$

where A is the semimajor axis of the transfer orbit. The cruise time is one-half of the period of the orbit, which is equal to $A^{1.5}$ (in years) if A is expressed in AUs). From Kepler’s laws, it is then easy to derive the heliocentric velocities at departure (V_1) and arrival (V_2):

$$V_1 = V_e(2 - 2/A)^{0.5} \quad \text{and} \quad V_2 = V_e(2/R_p - 2/A)^{0.5} \quad [5]$$

The orbital velocity V_p of a planet in a circular orbit is $V_e R_p^{-0.5}$. The relative velocity at departure (V_{dep}) and arrival (V_{arr}) is therefore

$$V_{\text{dep}} = V_1 - V_e \quad \text{and} \quad V_{\text{arr}} = V_2 - V_e R_p^{-0.5} \quad [6]$$

Table 2 Orbital elements of selected solar system bodies and main characteristics of the corresponding transfer orbits

Target	Inclination	Eccentricity	Semimajor axis (AU)	Departure (km s^{-1})	Arrival (km s^{-1})	Cruise (years)
Mercury	7.01°	0.206	0.387	-7.53	9.61	0.29
Venus	3.39°	0.007	0.723	-2.50	2.71	0.40
Mars	1.85°	0.093	1.524	2.94	-2.65	0.71
Vesta	7.13°	0.090	2.365	5.51	-4.43	1.09
Jupiter	1.31°	0.049	5.203	8.80	-5.64	2.73
Saturn	2.49°	0.056	9.539	9.60	-5.44	6.04

The relative velocities are evaluated assuming all planetary orbits to be circular, in the same plane as the orbit of the Earth. In this table, relative velocities are positive if the spacecraft goes faster than the planet or asteroid, negative if the spacecraft goes slower than the planet or asteroid. The departure and arrival conditions further assume that the Earth and the target body are at the appropriate relative positions in the solar system.

Table 2 presents the required departure and arrival velocities for selected solar system bodies in the circular and coplanar approximation. Mercury and Vesta depart very significantly from this hypothesis in terms of inclination (**Table 2**, column 2), and the eccentricity of the orbit of Mercury is large (**Table 2**, column 3). When comparing **Tables 2** to **1** and **Figure 1**, Venus and Mars stand out as targets which are relatively easy to reach, with optimum departure velocities in the range of $2.5\text{--}3 \text{ km s}^{-1}$ (respectively backward and forward relative to the orbital motion of the Earth at departure). Even a simple flyby to Vesta is already a challenge, as the departure velocity in this optimistic hypothesis is not far from the upper limit in **Figure 1**. Direct transfer missions to Mercury, Jupiter, Saturn, and even more so Uranus and Neptune are out of reach of medium-class launchers such as Soyuz if any significant mass is to be brought to destination. Another important conclusion of this preliminary analysis is that missions to Mars and Venus require relatively short transfer times, while missions to the outer solar system require relatively long cruise durations, apart from being energetically more demanding.

10.17.3.3 Rendezvous Missions

For a rendezvous mission, the spacecraft has to be captured in orbit around the target body. The simplest way to get into orbit is to brake closest to the target body using the spacecraft thruster. This is the exact counterpart of the launch strategy from a highly elliptical orbit which was considered in Section 10.17.3.2. The same issues arise on the length of thrust arcs, with a possible impact in terms of gravity losses if one has to thrust at higher altitudes.

The closest possible altitude for an orbit-insertion maneuver depends on the target body: if it has no atmosphere (Mercury, Vesta), 200 km is a safe distance. With an atmosphere (Mars, Venus), 300 km is a reasonable choice. If there are rings (Saturn) or intense radiation belts (Jupiter), larger distances may be required due to safety constraints, for example, 80 000 km for Saturn so as to be outside of the rings (one can get closer by flying above the rings, as was done by Cassini). At a given distance from a planet, the local escape velocity V_{esc} defines the orbits which have a total energy of 0 in the reference frame of the primary. The relationship between the ΔV required for a marginal capture and the arrival velocity V_{arr} can then be derived from energy conservation:

$$(V_{\text{esc}} + \Delta V)^2 = V_{\text{esc}}^2 + V_{\text{arr}}^2 \quad \text{and} \\ \Delta V = (V_{\text{esc}}^2 + V_{\text{arr}}^2)^{0.5} - V_{\text{esc}} \quad [7]$$

From eqn [7], the required ΔV increases with the arrival velocity and decreases with the local escape velocity, hence it is easier to capture around a large solar system body. The actual orbit-insertion maneuver is not much larger than that required for a marginal capture, as the first orbit around the target has always a much longer period than the operational orbit. The 8 days initial orbit of Mars Express costs 60 m s^{-1} more than a marginal capture around Mars, and the 115 days initial orbit of Cassini around Saturn costs 140 m s^{-1} more than a marginal capture around Saturn with the same approach velocity.

From **Table 3**, the large gravity fields of Jupiter and Saturn compensate for the relatively large arrival velocity (**Table 2**), so that the capture ΔV is in the same range as that required for Venus and Mars ($0.34\text{--}0.68 \text{ km s}^{-1}$). Conversely, orbiter missions

Table 3 Capture ΔV requirements for a rendezvous mission

Target	Arrival (km s ⁻¹)	Local escape velocity (km s ⁻¹)	Capture ΔV (km s ⁻¹)
Mercury	9.61	4.09	6.35
Venus	2.71	10.36	0.35
Mars	-2.65	4.81	0.68
Vesta	-4.43	0.3	4.14
Jupiter	-5.64	35.0	0.45
Saturn	-5.44	24.0	0.61

around Mercury or an asteroid such as Vesta require extremely large orbit-insertion maneuvers. Once again, it is useful to go back to the rocket equation [3]: with an exhaust velocity of 3.1 km s^{-1} , the mass in orbit can be larger than 80% of the mass during approach for ‘easy’ rendezvous targets such as Venus, Mars, Jupiter, and Saturn. Conversely, orbiter missions to Mercury and Vesta are out of reach of a single-stage onboard propulsion system unless the relative velocity can be reduced to more reasonable values using gravity-assist strategies (see Section 10.17.7).

10.17.3.3.1 Rendezvous missions to natural satellites of giant planets

Among the 15 largest solar system bodies (excluding the Kuiper Belt), seven are in orbit around a planet: the Moon and the six largest natural satellites of giant planets (Io, Europa, Ganymede and Callisto around Jupiter, Titan around Saturn, Triton around Neptune). It has already been indicated that a flyby mission to a Galilean satellite has very similar requirements to a flyby mission to Jupiter, and the Voyager 1 and 2 missions did combine these two objectives. The situation is drastically different for orbiter missions: the large gravity fields of giant planets, which make it relatively easy to capture into a highly elliptical orbit, result in very large orbital velocities and local escape velocities at the distances of large natural satellites. As an example, the orbital velocity of Europa is 13.7 km s^{-1} , and the local escape velocity is 19.4 km s^{-1} . Capturing a spacecraft in an orbit with its pericenter at the distance of Europa requires a minimum ΔV of 0.8 km s^{-1} (one can do somewhat better by capturing at a lower altitude then raising the pericenter, but this extends the mission duration). Matching the orbit of Europa for a rendezvous would require an additional 5.7 km s^{-1} . Therefore, orbiter missions to satellites of giant

planets cannot be achieved simply with a chemical propulsion system. Much more complex methods must be implemented, such as satellite gravity-assist methods (Longman and Schneide, 1970) or nuclear electric propulsion (NEP), or a combination of both (see Sections 10.17.7 and 10.17.8).

Apart from Mercury, all planets have a substantial atmosphere. Among natural satellites, this is the case only for Titan. It is then possible to dip into the atmosphere so as to lose energy and lower the orbit (French and Cruz, 1980). This has to be done in one pass if one wants to use the atmosphere for the insertion maneuver (aerocapture) or for a landing module (atmospheric entry). Once in orbit, it is possible to lower the orbital period by dipping into the outer regions of the atmosphere during many successive orbits (aerobraking). *In situ* missions to planetary bodies with an atmosphere have a close relationship with flyby missions, as an entry module with a heat shield can dissipate quite effectively the arrival energy for a wide range of arrival velocities.

10.17.3.3.2 Rendezvous missions to comets and Earth-crossing asteroids

Comets and Earth-crossing asteroids are very specific, as flyby missions and rendezvous missions have completely different characteristics, contrarily to solar system bodies in nearly circular orbits. As has already been mentioned, flyby missions can be very cheap if the encounter location is not too far from the orbit of the Earth. Unfortunately, such encounters occur at a very high relative velocity (more than 70 km s^{-1} for the flybys of Halley’s comet in 1986), which make it impossible to consider a rendezvous. Two conditions must be met for a rendezvous mission to be feasible with chemical propulsion:

1. the inclination must be relatively small ($\leq 12^\circ$), which eliminates most candidates and
2. the perihelion must range between 0.65 and 1.5 AU.

Asteroid 433 Eros, the largest near-Earth asteroid, meets these criteria with a perihelion at 1.13 AU and an inclination of 10.8° . It was studied by the Near Earth Rendezvous mission of NASA (NEAR) from rendezvous in Feb 2000 to a programmed touch-down on Eros in Feb 2001 (Cheng *et al.*, 1997). Comet Churyumov-Gerasimenko, the target of the Rosetta rendezvous mission, is a typical good candidate among comets, with an inclination of 7.1° and a perihelion at 1.29 AU.

The lowest relative velocity at arrival is achieved by departing from the Earth close to the perihelion of the target body, hence arriving close to aphelion. Again considering the case of Churyumov-Gerasimenko, the minimum arrival velocity is a reasonable 1.2 km s^{-1} , but the departure velocity is close to 10 km s^{-1} , which is out of reach for a direct launch. Similarly to giant planets, gravity-assist strategies (Section 10.17.7) or an advanced propulsion system (Section 10.17.8) are required.

10.17.4 Remote-Sensing and *In Situ* Missions to Venus and Mars

The simplified analysis in Section 10.17.3, demonstrated that Venus and Mars stand out as the easiest targets for a rendezvous mission apart from the Moon. They are also surrounded by an atmosphere, which is a major asset for implementing *in situ* modules (landers, rovers, balloons, etc.). It is therefore no surprise that until 1972 all planetary missions were sent towards one of these two planets. However, not all opportunities are equal, due to the non-negligible inclination of Venus (3.39°) and Mars (1.85°) as well as the significant eccentricity of the orbit of Mars ($e=0.093$).

The first issue to consider is how often can one send a spacecraft to Venus and Mars. Arrival occurs nearly 180° away from departure. During the required cruise time from Earth to Venus on a Hohmann transfer trajectory (0.399 years), Venus travels 233° . Therefore, the appropriate departure time is when Venus is 53° behind the Earth on its orbit. For Mars, the cruise time is 0.709 years, during which Mars travels 135.6° . The appropriate departure time is when Mars is 44.4° ahead of the Earth on its orbit (this assumes that the orbit of Mars is nearly circular, which is not the case). How often does such a favorable configuration repeat? This period P_{syn} (called the synodic period) is such that the angle traveled by the innermost planet (which moves fastest) exceeds by one full turn the angle traveled by the outermost planet. If P_1 and P_2 are the periods of the innermost and outermost planet, we have

$$P_{\text{syn}}/P_1 = P_{\text{syn}}/P_2 + 1, \quad P_{\text{syn}} = P_1 P_2 / (P_2 - P_1) \quad [8]$$

If one considers the orbital period of Venus (0.615 years) and Mars (1.881 years), this gives a repeat interval of 1.6 years for transfers to Venus and 2.135 years for transfers to Mars. The position of the Earth at launch moves forward by 216° for Venus and by

48.6° for Mars. After 8 years (five synodic periods, 13 Venus ‘years’), the windows to Venus repeat almost perfectly, as Venus is then only 1.5° ahead of its position 8 years before. For Mars, the best match is after 15 years (seven synodic periods), but Mars lags by 9° behind its position 15 years before, and the windows do not repeat. Intermediate launch dates can be obtained by inserting ‘phasing’ orbits in the transfer, so that the spacecraft travels close to 540° instead of 180° . This increases the cruise time by 0.8 years for Venus and 1.4 years for Mars. A 540° transfer to Venus requires launching 6 months earlier than the direct window. Such opportunities can be useful, in particular when Venus or Mars is used for a gravity assist (see Section 10.17.8).

The significant eccentricity of the orbit of Mars (0.093) has an impact on transfer orbits. The minimum velocity at infinity required to reach Mars is 3.5 km s^{-1} when it is reached at aphelion (1.665 AU) and only 2.3 km s^{-1} when it is reached at perihelion (1.38 AU). Conversely, the minimum arrival velocity is much higher at perihelion (3.3 km s^{-1}) than at aphelion (2.1 km s^{-1}). The range of departure velocities corresponds to a difference of only 300 m s^{-1} of ΔV from the upper stage. The range of arrival velocities corresponds to a difference of 640 m s^{-1} for the insertion maneuver (actually somewhat larger as gravity losses also increase), which is quite significant (20% less mass budget for an orbiter with an insertion at perihelion, if one does not use aerocapture or aerobraking). Venus has a nearly circular orbit, so that in this respect all Venus windows should be equivalent. The Earth itself has a small eccentricity (0.0167), so that an outbound launch in Jan (when the Earth is closest to the Sun) and an inbound launch in Jul require slightly smaller departure velocities.

However, a much more important problem results from the small but significant inclination of the orbit of Mars (1.83°) and even more so Venus (3.39°) with respect to the orbit plane of the Earth (the ecliptic). The plane of the transfer orbit contains the Sun, the Earth at departure, and the target planet at arrival. In most cases, the planet at arrival is not in the ecliptic. This only happens when it lies at one of the two intersections of the orbit plane and the ecliptic: $L_s \sim 76.8^\circ$ and $L_s \sim 256.8^\circ$ for Venus, $L_s \sim 49.5^\circ$ and $L_s \sim 229.5$ for Mars, where L_s is the heliocentric longitude, measured from the Sun–Earth direction at the fall equinox. This forbids a perfect Hohmann transfer, as a 180° orbit would have to fly directly over the Sun. For Mars, angles of 160° (type I transfers) or 200° (type II transfers) bring the transfer orbit plane

much closer to the ecliptic. The in-plane departure and arrival velocities are suboptimal compared to the Hohmann transfer, but the out-of-ecliptic component of the departure velocity drops from the staggering values ($>30 \text{ km s}^{-1}$) required for a 180° transfer to 2 km s^{-1} or less (when Mars at arrival is close to one of the nodes). For Venus, with its larger inclination, angles of 135° (type I) and 225° (type II) between departure and arrival have to be used so as to reduce the out-of-ecliptic component of the departure velocity to 2 km s^{-1} or less.

As a conclusion of this analysis, a good window to Mars (Venus) occurs when Mars (Venus) can be reached close to one of the nodes. For Venus, both nodes are equivalent, as they are at nearly the same distance from the Sun (0.7205 and 0.7261 AU for L_s 76.8° and L_s 256.8° , respectively), but only the latter lies conveniently close to one of the five possible arrival locations in the 8-year cycle. For Mars, arriving close to L_s 49.5° (at 1.4715 AU) is slightly better than arriving close to L_s 229.5° (at 1.5515 AU) for flyby and *in situ* missions and slightly worse for orbiter missions relying only on chemical propulsion.

The situation is further complicated as the performances of launchers from Kourou (Ariane 5, Soyuz–Fregat starting in 2009) depend on the declination (angle of the departure velocity to the equatorial plane of the Earth). At the low latitude of Kourou (5° N), the launcher can take full advantage of the rotational velocity of the Earth (461 m s^{-1} instead of 463 m s^{-1} at the equator) if the departure velocity lies close to the equatorial plane. For a given departure velocity, the best windows are those with a low departure declination. When the launch pad is at a higher latitude (Kennedy Space Center: 28.5° N ; Baikonur: 46° N), the rotational velocity is lower (407 m s^{-1} for Kennedy, 322 m s^{-1} for Baikonur), but the launch capability is the same for all declinations (N or S) which do not exceed the latitude, as a restartable upper stage (Centaur, Fregat) is required for such launch sites. As there is an angle of 23.45° between the ecliptic and the equator, the declination is a complex combination of the departure date and the out-of-ecliptic component.

Tables 4a and **4b** provide useful comparisons for all remote-sensing and *in situ* missions to Mars and

Table 4a Launch windows to Venus for Venus Express (10/2005) and from 2013 to 2018

Launch	V_{dep} (km s^{-1})	Launch declin.	$\Delta V/\text{Moon}$	Arrival position	Arrival date	V_{arr} (km s^{-1})	$\Delta V/V_{\text{esc}}$	ΔV Total	Mass (kg)
Coplanar appr.	2.500		0.338			2.710	0.357	0.695	1287
2005/10/31 (II)	2.780	26.4° S	0.404	$L_s 249^\circ$	2006/04/07	4.653	1.019	1.423	1017
2013/10/30 (II)	2.795	26.6° S	0.408	$L_s 250^\circ$	2014/04/06	4.730	1.052	1.459	1006
2015/05/21 (II)	2.480	9.7° N	0.333	$L_s 78^\circ$	2015/10/27	3.780	0.683	1.017	1160
2016/12/28 (I)	2.664	3.0° N	0.376	$L_s 257^\circ$	2017/05/09	3.754	0.674	1.050	1147
2018/06/15 (II)	3.891	-42.2° S	0.728	$L_s 116^\circ$	2018/12/17	3.061	0.453	1.182	1100
2018/08/22 (I)	3.107	9.1° N	0.489	$L_s 102^\circ$	2018/12/08	4.562	0.981	1.470	1002

The type of transfer (I or II) is indicated. The ΔV at departure is evaluated assuming an indirect launch from a Moon-crossing orbit. The ΔV at arrival is evaluated for a marginal capture at an altitude of 300 km. The mass in orbit is evaluated for a Venus Express-type launch (Soyuz/Fregat from Baikonur: no penalty for declination up to 46° N or S).

Table 4b Launch windows to Mars for Mars Express (06/2003) and from 2005 to 2018, with the same definitions as for **Table 4a**

Launch	V_{dep} (km s^{-1})	Launch declin.	$\Delta V/\text{Moon}$	Arrival position	Arrival date	V_{arr} (km s^{-1})	$\Delta V/V_{\text{esc}}$	ΔV Total	Mass (kg)
Coplanar appr.	2.940		0.444			2.650	0.681	1.125	1120
2003/06/03 (I)	2.990	3.9° S	0.458	$L_s 50^\circ$	2003/26/12	2.693	0.702	1.159	1108
2005/08/23 (II)	4.000	0.3° N	0.766	$L_s 183^\circ$	2006/08/26	3.102	0.912	1.678	937
2007/09/14 (II)	3.638	9.1° N	0.646	$L_s 203^\circ$	2008/08/26	2.531	0.625	1.270	1069
2009/10/14 (II)	3.200	20.6° N	0.515	$L_s 228^\circ$	2010/09/03	2.465	0.594	1.109	1126
2011/11/11 (II)	3.000	32.3° N	0.460	$L_s 258^\circ$	2012/09/17	2.720	0.715	1.175	1102
2013/12/06 (II)	3.075	24.5° N	0.480	$L_s 289^\circ$	2014/09/27	3.180	0.955	1.435	1013
2016/03/07 (I)	3.355	46.1° S	0.559	$L_s 314^\circ$	2016/09/24	3.855	1.353	1.912	869
2018/05/13 (I)	2.800	35.2° S	0.409	$L_s 26^\circ$	2018/12/05	2.945	0.829	1.238	1080

Venus. They show the major impact of the 3.39° inclination of Venus on the approach velocity, which never gets close to the lower limit derived from the coplanar assumption. Conversely, several Mars windows have overall performances which are very close to that of the coplanar assumption. The departure or arrival velocity for Mars can even be smaller than the coplanar estimate, as Mars (0.093) and the Earth itself (0.0167) have noncircular orbits. The similarity between the Venus Express transfer orbit (2005/10) and that of the 2013 window illustrates the 8-year cycle of windows to Venus.

10.17.4.1 Orbiter Missions Not Implementing Aerocapture

Orbiter missions that do not implement aerocapture have to use their propulsion system to insert into orbit around the planet. The simplest way to compare windows is to consider as a baseline for all missions the launch strategy from a Moon-crossing orbit described in Section 10.17.3.2 which provides the best mass budget on Mars and Venus transfer trajectories. The spacecraft propulsion system is used both to leave Earth orbit and for orbit insertion. Therefore, high departure or arrival velocities have a comparable impact in terms of total ΔV (next to last column of Tables 4a and 4b). Assuming a 3.1 km s⁻¹ ejection velocity for the spacecraft thruster, this can be translated into an available mass in orbit by applying the rocket equation [3] to the launch capability in a large bound orbit for a given launcher (Soyuz–Fregat for the last column in Tables 4a and 4b).

After capture in a long-period orbit (8 days for Mars Express), it is possible to reduce the period of the orbit by implementing thrust at pericenter. Additional ΔV (2.96 km s⁻¹) are required to go down to a circular orbit around Venus at an altitude of 300 km, and 1.41 km s⁻¹ of additional ΔV are required to reach a circular orbit around Mars at an altitude of 300 km. Splitting the maneuver over many pericenters (eight for Mars Express) limits gravity losses to a few meters per second, as each thrust arc is short, close to the optimum altitude.

From Table 4a, the 2005 window to Venus was attractive in terms of departure velocity, but it was relatively poor in terms of arrival velocity. This was not a major issue as the orbital period could not be shorter than 1 day due to constraints on eclipse duration imposed by the reuse of the Mars Express spacecraft. The windows to Venus in mid-2015, late 2016, and mid-2018 are favorable, which is an asset

for proposals to ESA in the 2015–20 time frame ('Cosmic Vision').

From Table 4b, the 2003 window to Mars stands out as one of the best over 20 years. With a direct injection by Soyuz–Fregat, the Mars Express spacecraft (1050 kg after releasing the 70 kg Beagle 2 module) could brake down to a 6.7 h polar orbit, which required 1.5 km s⁻¹ of ΔV including a plane turn, leaving 70 kg of reserve fuel in the tanks on top of the 570 kg dry mass. The 2005 window is among those delivering the lowest mass in Mars orbit (20% less than in 2003, representing, e.g., 170 kg of fuel). This was one of the arguments against flying a new mission to Mars as a follow-up to Mars Express, selecting instead Venus Express which was successfully launched in Oct 2005. The recently approved exploration program of ESA is considering a mission in the 2011–15 time frame. As the present design is based on a large rover ('ExoMars'), this mission belongs to the category of *in situ* missions, which will be considered next. Orbiter missions may be required after 2010 so as to provide relay capability. In this respect, 2013 is poor and 2016 is bad, and the next good window after 2011 is in 2018.

The mass budget of orbiter missions can be markedly improved by aerobraking. This method involves shallow atmospheric passes which slowly reduce the semimajor axis starting from a highly elliptical capture orbit. The impact on spacecraft design is significant, but far less severe than for aerocapture. Solar panels pick up most of the strain. They have to be symmetrical and relatively robust, but no heat shield is required. This method was first implemented for the MGS mission of 1998 (Albee *et al.*, 2001) then for Mars Odyssey in 2000 (Smith and Bell, 2001), and in late 2006 for the MRO mission which was launched in Aug 2005. Aerobraking was definitely required for MRO as it is a large spacecraft while the 2005 window is poor. If one goes all the way down to a circular orbit, which is the case both for MGS and MRO, nearly 1.4 km s⁻¹ of ΔV are saved, which corresponds to a factor of 1.57 in terms of mass budget (actually closer to 1.7 if one takes into account the additional mass of tanks and structure). There is a drawback: as each pass has to be gentle, the orbit reduction takes a long time. The science phase of the mission was therefore delayed by 6 months for MRO, and it was delayed by 1 year for MGS, due to a non-nominal configuration of one of the solar panels. There is some risk involved, and the aerobraking phase requires a high level of monitoring, with a significant impact on operational cost. Aerobraking

has yet to be implemented by ESA, but it should definitely be considered for large orbiters in low orbits around Mars or Venus.

10.17.4.2 *In Situ* Missions and Orbiter Missions Implementing Aerocapture

For *in situ* missions (landers, rovers, balloons, etc.) or orbiter missions implementing aerocapture, an entry module with a heat shield dissipates the approach energy. The control of a trajectory deep in an atmosphere is a complex problem which is beyond the scope of this chapter. It is extensively discussed for Earth re-entry trajectories (see, e.g., Willcockson, 1999; Wu *et al.*, 2001) as well as for other planetary bodies (see, e.g., Scoon *et al.*, 1991; Justus *et al.*, 2005). There is a narrow entry corridor for aerocapture: if the minimum altitude is too high, the velocity remains above escape, and the spacecraft leaves the planet. If the minimum altitude is too low, the velocity drops below the circular velocity, and the spacecraft falls on the planet. The latter case is what is required for an *in situ* mission. However, if the entry is too steep, the heat flux may overcome the dissipation capability of the shield, and the mission is lost. The energy corresponding to the escape velocity at the surface (10.3 km s^{-1} for Venus, 5.1 km s^{-1} for Mars) has to be dissipated anyway, hence the mass which is required for the heat shield depends only weakly on the approach velocity: the range of specific energies to be dissipated is $57\text{--}64 \text{ MJ kg}^{-1}$ for Venus (see, e.g., Craig and Lyne, 2005), $16.5\text{--}20.5 \text{ MJ kg}^{-1}$ for Mars for the full range of approach velocities in **Tables 4a** and **4b**. Therefore, for such missions, the most important parameters are the departure performances, which can be defined from columns 3 (declination) and 4 (ΔV at departure from a bound orbit) of **Tables 4a** and **4b**. For such types of missions, a window with a low departure velocity and a high arrival velocity becomes attractive, and the type I window to Venus in 2018 would be selected (3.1 km s^{-1} at departure) instead of the type II window (3.7 km s^{-1} at departure) while the later would provide more mass for an orbiter with chemical insertion. ESA has demonstrated its capability to deliver a surface station to Titan with the Huygens mission, with an entry velocity of 6 km s^{-1} very similar to Mars entry velocities. In 2011 and 2013, a Soyuz–Fregat launched from Kourou can deliver a 1300 kg entry module to Mars. Even larger masses can be delivered to Venus in 2015 and 2016. Typically half of the mass of an entry module is

available as useful mass on the surface, as demonstrated by missions such as Huygens (ESA) or the 2004 Mars rovers (NASA). Therefore, a relatively large rover can be delivered to Mars with Soyuz/Fregat in 2011 and 2013, as presently considered for the Exomars mission of ESA's exploration program. Windows involving an additional orbit around the Sun are presently considered, so as to land well in advance of the southern summer dust storms. More capable launchers such as Ariane 5 (3500 kg delivered to Mars), would make it possible to implement a larger rover and a relay orbiter, but the mission would be correspondingly more expensive. The MSL mission of NASA, to be launched in 2009, is an example of such ambitious undertakings. Similarly to Venus and Mars, aerocapture provides the most effective strategy for other planetary bodies with a significant atmosphere. It is in particular likely that a post-Cassini mission to Titan will have to implement such a strategy (e.g., Spilker, 2005)

10.17.5 Orbit Evolution around Planetary Bodies

For all orbiter missions, the evolution of the orbit with time is a critical aspect of science operations, as it defines which regions can be observed as well as the local time (a critical parameter for cameras and near-IR spectrometers, which rely on solar photons).

The perturbations of the orbit by the Sun play only a minor role except for highly eccentric orbits. The main effect is a change in the minimum altitude, which has to be compensated if it would result in a crash. When approaching a planet at a low velocity, the combined gravitational pull of the Sun and the planet can provide for free a capture in a weakly bound orbit. These three body effects are important for planetary systems, in particular for the Earth–Moon system (Belbruno and Miller, 1993; Knezevic and Milani, 1998).

High-order terms of the gravity field of the planetary body are most important for orbits closest to the target body. They result in a precession of the intersection of the orbit plane with the equatorial plane (defined by the position of the ascending node Ω) as well as a precession of the pericenter (defined by the angular distance from the ascending node to the pericenter ω) for elliptical orbits. The quadrupolar term of the gravity field, \mathcal{J}_2 , plays the most important role. It is linked to the oblateness of

the planetary body. The evolution of Ω and ω as a function of \mathcal{J}_2 and orbital elements is as follows:

$$\dot{\Omega} = -\frac{3}{2}n\mathcal{J}_2 \left(\frac{R_m}{p}\right)^2 \cos i \quad [9]$$

$$p = a(1-e^2)$$

$$\dot{\omega} = -\frac{3}{4}n\mathcal{J}_2 \left(\frac{R_m}{p}\right)^2 (4-5\sin^2 i) \quad [10]$$

R_m is the radius of the planetary body, a is the semimajor axis, e is the eccentricity, n is the mean motion of the spacecraft on its orbit (inversely proportional to $a^{1.5}$), and i is the inclination. For circular orbits, the parameter (p in eqn [9]) is equal to a , and the precession of the plane is inversely proportional to $a^{3.5}$.

A specific case of interest is the sun-synchronous orbit, with a full rotation of the orbit plane in exactly one orbital period of the planetary body around the Sun. With such an orbit, the local time remains approximately constant. This is a desirable feature for cameras and IR imaging spectrometers as observations on successive orbits are made with the same illumination conditions. With the value of \mathcal{J}_2 for the Earth (0.001082), an inclination of 97.8° is needed for a circular orbit at 600 km altitude from eqn [9]. A sun-synchronous orbit has been selected for NASA missions orbiting Mars on a low circular orbit, such as MGS and MRO. The \mathcal{J}_2 value for Mars (0.001965) is larger than that of the Earth and the orbital period of Mars is 687 days, so that a lower value of $\cos(i)$ is required for sun-synchronous orbits. The inclination of MGS (altitude ~400 km) is 92.9° and that of MRO

(which reached a circular orbit at an altitude of ~300 km in Nov 2006) is 92.7°. MRO crosses the equator in mid-afternoon (3 p.m.) on average. However, the rate of precession is constant, while the angular motion of Mars is 45% larger at perihelion (1.382 AU) than at aphelion (1.666 AU) due to Kepler's second law. As a consequence, the local time at the equator varies from 2.15 to 3.45 p.m. during one Martian year. The variations in local time are only 2 min for sun-synchronous orbits around the Earth due to the low eccentricity of its orbit (0.0167).

The precession of the pericenter of an elliptical orbit around Mars such as that of Mars Express is defined by eqn [10]. This precession provides opportunities for observing each latitude range at the highest possible resolution. However, combining the rotation of the planet around the Sun, the precession of the orbit plane and the precession of the pericenter so as to optimize the science return is a complex problem (Hechler *et al.*, 2005), in particular when both dayside observations (for imaging experiments) and nightside observations (for the MARSIS radar) are required at low altitude. The orbit evolution of the selected orbit is presented in Figure 2. It provides deep excursions of the pericenter over the sunlit side (L_s 135°, L_s 315°, L_s 90° in 2006) as well as shallow excursions of the pericenter beyond the terminator (L_s 90°, L_s 225°)

Venus has a very small oblateness, which can be attributed to its very slow rotation rate. Therefore, precession was neither an issue nor an asset for orbiter missions to Venus such as Magellan (NASA) or Venus Express (ESA). The pericenter of the highly

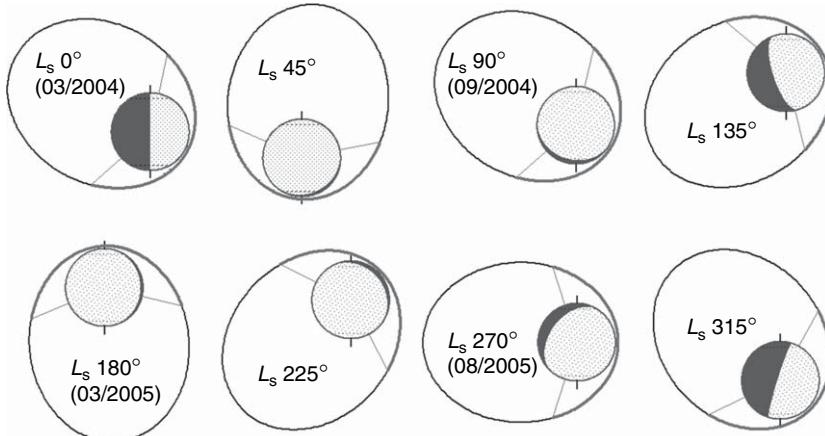


Figure 2 Evolution of the orbit of Mars Express during the first Martian year of science operations. For each heliocentric longitude L_s , the viewing direction is perpendicular to the orbit. The pericenter rotates 1.22 times around the planet during one Martian year, so that each latitude is observed at least twice at the lowest altitude (~280 km).

elliptical orbit of Venus Express will therefore remain at high northern latitudes for the full science mission. The three Mariner 10 flybys left significant uncertainties on the high-order terms of the gravity field of Mercury. The most likely value of \mathcal{J}_2 is very small (0.00008). The resulting precession is negligible for the highly eccentric orbit of MESSENGER (NASA). One of the two orbiters of the Bepi-Colombo mission (ESA/JAXA) has a low eccentricity (0.16), and the precession of the pericenter could reach $40^\circ\text{--}50^\circ$ over the 2-year-long observation phase. It is therefore planned to insert the spacecraft with a pericenter at 20° S , then let it drift northward so as to optimize observation conditions.

10.17.6 *In Situ* Missions to Atmosphereless Bodies

In situ missions to atmosphereless bodies represent a specific challenge, as one cannot rely on an atmospheric entry module. As indicated in Section 10.17.3, such modules are quite effective as the escape energy and the approach energy can be dissipated in the atmosphere, with a useful mass which is about half of that of the entry module. There is always an element of risk (failure of Mars Polar Lander in 1999 and Beagle in 2003), but the examples of the 2004 Mars rovers (NASA) or Huygens (ESA) show that this technology is now mature. On atmosphereless bodies, the only solution for landing safely on the surface consists in using an onboard propulsion system. A soft landing is in many ways similar to a launch: it requires that the acceleration from the thrusters exceeds the acceleration from surface gravity close to the surface. There is one piece of good news: contrarily to a launch, the mass decreases as the landing module gets near the surface, hence the acceleration from the thrusters increases close to the surface, while the acceleration from gravity remains nearly constant. Again similarly to a launch, a landing requires a ΔV which is larger than the escape velocity (or the orbital velocity at pericenter if the lander is released from an orbiter). The

propulsion system has to land itself, and the mass penalty for very large thrust levels needs to be considered. Conversely, if the thrust is too low, there are gravity losses (part of the thrust is performed at large distances from the planet) and a penalty in the final stages when part of the thrust is wasted in fighting the local gravity.

Atmosphereless solar system bodies of interest are comets, asteroids, KBOs (among them Pluto and probably Triton), small natural satellites, six of the seven largest natural satellites (the Moon, Io, Europa, Ganymede, Callisto, and Triton), and one planet (Mercury). Landing on a cometary nucleus a few kilometers across is a difficult navigation problem but it is not a mass-budget issue, as escape velocities are only in the range of meters per second. There is no survey mission planned yet for Pluto or Triton, therefore an *in situ* mission to KBOs is far in the future. The most relevant issues are therefore that of landing on Mercury, the Moon, Galilean satellites, and large asteroids such as Vesta. It may come as a surprise to list the Moon among relevant targets, as there are already several 100 kg of samples from our satellite, but high latitudes and the farside remain virgin territory for *in situ* missions.

Table 5 shows the escape velocity and surface gravity of relevant large atmosphereless bodies. From **Table 5**, it is clear that in terms of total ΔV there is not much difference between a soft lander (a few meters per second), a semihard lander with airbags ($\sim 50\text{ m s}^{-1}$), and a penetrator (up to 100 m s^{-1}), with the exception of asteroids such as Vesta or Ceres.

Table 5 shows that Mercury, while not the largest atmosphereless body, is in a class by itself in terms of landing: the surface gravity is more than twice that of any other atmosphereless body, and the energy to be dissipated is 2.35 to 4 times larger than that required for Galilean satellites. The mass budget required for a lander is therefore very large.

Ion propulsion missions (see Section 10.17.8) result in an approach to Mercury on a weakly bound orbit. The key issue is the fraction of the mass approaching Mercury which can reach the surface and the smaller fraction which is left for lander

Table 5 Escape velocity and surface gravity of large atmosphereless bodies

	Mercury	Moon	Io	Europa	Ganymede	Callisto	Vesta	Ceres
Size (km)	4880	3470	3640	3120	5280	4820	540	945
V_{esc} (km s^{-1})	4.257	2.376	2.560	2.030	2.741	2.440	0.370	0.515
g (m s^{-2})	3.700	1.622	1.800	1.320	1.430	1.235	0.260	0.280

systems and payload ('useful mass') once the mass of the propulsion system (thrusters, tanks, structure) has been accounted for. The critical parameter is the acceleration level at approach. This defines the length of the arc which is required to brake into orbit, then to land. The longer the thrust arc, the farther from Mercury it extends, with corresponding gravity losses. It is simpler not to split the insertion between several successive orbits (which would save a few kilograms) and to continue thrusting until landing. With this approach, the propulsion system does not need to be restarted, which constitutes an asset in terms of risk.

As soon as the spacecraft velocity is lower than that of a circular orbit, the centrifugal force is lower than the local gravity. A simple landing strategy, which is close enough to optimum for this order-of-magnitude analysis, consists in tilting the thrust towards the center of the planet so that the vertical component balances gravity together with the remaining centrifugal force. The remaining horizontal component is used to reduce the orbital velocity. The minimum total ΔV expenditure from a weakly bound orbit (4.24 km s^{-1}) is close to the escape velocity (4.26 km s^{-1}) if and only if the velocity close to the surface can be killed by a single impulse. This defines the upper limit of the mass on the surface, requiring unrealistically high thrust levels. **Table 6** indicates the thrust time, total ΔV , and mass breakdown of the landing module for acceleration levels between 0.75 and 2 m s^{-2} .

As expected, a larger thrust level provides a larger landing mass as less time is spent fighting the 3.7 – 3.8 m s^{-2} of local gravity. However, the mass of the propulsion system must be taken into account. If we consider a thruster mass of 15 kg for 400 N of thrust, one needs 3.75% of the initial mass ($15 \text{ kg}/400 \text{ kg}$) for a 1 m s^{-2} initial acceleration. Ten percent of the fuel mass (itself up to 89% of the initial mass, in the 0.75

m s^{-2} case) is typically required for tanks and their support structure. Once these two lines have been subtracted, the useful mass on the surface turns out to be either negative or very low with a single-stage system for initial acceleration levels lower than 1.5 m s^{-2} . This useful mass must be adequate for a lander which is capable of driving the composite to the surface (i.e., with some simple altitude control system). Therefore, a useful mass of 50 kg for 5 – 10 kg of science payload can be considered as a minimum. With an initial acceleration level of 1.5 m s^{-2} , this useful mass has to be multiplied by ~ 16 ($1/0.064$) to obtain the initial mass. Therefore, a minimum mass of 800 kg (margins excluded) is required for the landing module approaching Mercury, 160 kg reaching the surface, with a thrust level of 1200 N . The results of this very simplified analysis are fully consistent with that of the ESA study for the BepiColombo mission to Mercury. Such a huge landing module required by itself a full Soyuz-Fregat, as a dedicated solar electric propulsion (SEP) stage (500 kg), cruise module (150 kg), and xenon fuel (400 kg) were needed. These technical elements played a major role in the decision to descope the landing module in mid-2003, when compatibility with a cornerstone budget became an issue.

Other large atmosphereless bodies provide much more favorable constraints for a lander. It turns out that a 1 m s^{-2} acceleration level at approach is adequate for all these bodies. **Table 7** provides the minimum mass of a landing module for a 50 kg smart lander which controls its trajectory to the ground.

The important role of surface gravity is demonstrated by the slightly more favorable mass budget for Ganymede than for Io, while its escape velocity is significantly larger (2.74 km s^{-1} instead of 2.65 km s^{-1} , **Table 5**). The large size of Ganymede also helps as thrust arcs are smaller when compared

Table 6 Elements of a mass budget tradeoff for a Mercury lander resulting from the simple landing strategy presented in the text

Initial acceleration	0.75 m s^{-2}	1 m s^{-2}	1.5 m s^{-2}	2 m s^{-2}	∞
Thrust time (s)	3660	2640	1650	1200	0
Total $\Delta V (\text{km s}^{-1})$	6.70	5.88	4.99	4.62	4.24
Mass on the surface	0.115	0.150	0.200	0.227	0.255
Propulsion system	0.028	0.038	0.056	0.075	NA
Tanks and structure	0.089	0.085	0.080	0.077	NA
Useful mass	−0.002	0.026	0.064	0.075	NA

The total mass on the surface, the mass of the propulsion system, the mass of tanks and structure, and the useful mass are expressed as a fraction of the mass of the landing module at approach.

Table 7 Mass of the landing module on a low velocity approach to four medium-sized atmosphereless bodies of interest which is needed to implement a smart lander with a useful mass of 50 kg, assuming a 1 m s^{-2} acceleration level at approach on a weakly bound orbit

	Moon	Io	Europa	Ganymede
Impulsive $\Delta V (\text{km s}^{-1})$	2.37	2.57	2.03	2.75
Actual $\Delta V (\text{km s}^{-1})$	2.93	3.17	2.40	3.12
Mass on the surface	0.387	0.359	0.460	0.365
Propulsion system	0.038	0.038	0.038	0.038
Tanks and structure	0.061	0.064	0.054	0.063
Useful mass	0.289	0.257	0.368	0.264
Mass of landing module (kg)	173	194	136	189

Similarly to [Table 6](#), masses are expressed as a fraction of the initial mass on such an orbit.

to the planetary radius, which reduces gravity losses. Even with more conservative assumptions, it is clear that a landing module on these bodies (in particular Europa) can be considered for a mass budget not exceeding 200–250 kg at approach. This mass budget has of course to be multiplied by the proper factor from the rocket equation [3] to account for ΔV expenditures from departure to approach at a low velocity.

10.17.7 Gravity-Assist Missions: Giant Planets, Mercury, Asteroid, and Comet Rendezvous

As demonstrated in [Section 10.17.3.2](#), flyby missions to giant planets and Mercury as well as rendezvous missions to comets require departure velocities

higher than 7 km s^{-1} , which are difficult to reach with large useful masses by a direct launch or an indirect launch strategy. A wide range of missions to these bodies became possible by using gravity-assist strategies.

A planetary gravity-assist consists in passing close to a planet so as to change the trajectory of a spacecraft. In the planetary referential, energy must be conserved, hence the relative velocity at departure is the same as the relative velocity at arrival. However, the planetary encounter can rotate this relative velocity by an angle which increases with the mass of the planet and decreases with the distance at closest approach. The benefit appears in the heliocentric referential: passing behind a planet significantly increases the heliocentric velocity ([Figure 3](#)). In such a case, the planet has provided

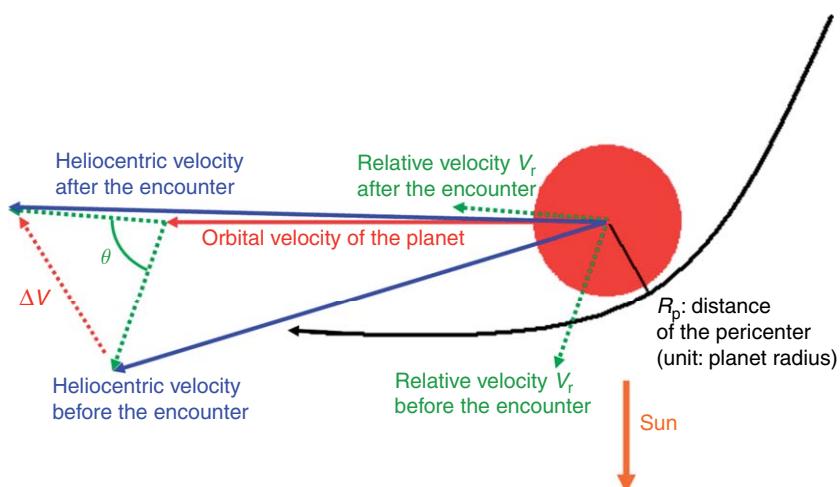


Figure 3 Example of a planetary gravity-assist which increases the orbital energy of a spacecraft in the heliocentric referential. The spacecraft passes over the trailing side. The gravitational pull of the planet rotates the relative velocity at arrival (mostly inbound) by an angle θ in the forward direction at departure, providing a ‘free’ ΔV . The resulting heliocentric velocity is therefore larger than at arrival. Such an encounter is included in any DVGA strategy.

additional orbital energy to the spacecraft in the heliocentric referential. Similarly, passing in front of the planet reduces the orbital energy without expenditure of onboard fuel.

The geometry of the swingby is defined by the relative velocity at arrival V_{rel} and the minimum approach distance R_p compared to the radius R of the planetary body and the circular velocity V_c at the surface (a factor of $\sqrt{2}$ smaller than the escape velocity V_{esc}). V_c is directly linked to the gravitational potential of the planet GM . The determination of the semimajor axis (a) and eccentricity (e) of the hyperbolic trajectory around the planet is straightforward from Newton's equations of motion:

$$\begin{aligned} GM/R^2 &= V_c^2/R \\ a &= -GM/V_r^2 = R(V_c/V_{\text{rel}})^2 \\ e &= 1 - R_p/a \end{aligned}$$

The true anomaly ν of the asymptote is defined by

$$1 + e \cos(\nu) = 0$$

The angle of rotation θ is:

$$\theta = 2\nu - \pi$$

The relationship between the rotation angle θ , the minimum distance R_p , and the relative velocity V_r can then be established, as well as the change in velocity ΔV :

$$\theta = 2a \sin(1/e) = 2a \sin(1/(1 + R_p/R(V_{\text{rel}}/V_c)^2)) \quad [11]$$

$$\Delta V = 2V_{\text{rel}}/(1 + R_p/R(V_{\text{rel}}/V_c)^2) \quad [12]$$

From eqn [12], the maximum theoretical ΔV provided by the planet is equal to V_c . It is obtained for $V_{\text{rel}} \sim V_c$ and $R \sim R_p$. From **Table 8**, any planetary encounter can provide a large ΔV for free. This is also the case to a lesser extent for large natural satellites (**Table 5**).

The first approach which can be considered is the ΔV gravity assist (DVGA) which increases the effectiveness of the available onboard ΔV by a large factor by combining a deep space maneuver with a

gravity-assist by the Earth (see, e.g., Casalino *et al.*, 1998). DVGA strategies can also be used with other planets or natural satellites. A DVGA strategy can be associated to any resonant orbit which goes back to the Earth after a given number of revolutions. The departure velocity is parallel to the Earth velocity (either forward or backward). The deep space maneuver is performed close to aphelion, reducing the semimajor axis for orbits with periods larger than 1 year, close to perihelion, increasing the semimajor axis for orbits with periods smaller than 1 year. As a result of this maneuver, the relative velocity when returning to the Earth increases by a much larger amount than the ΔV spent in the maneuver. This is due to the relationship between energy (E) and angular momentum (M) at pericenter ($p = 1 \text{ AU}$) and apocenter (q). The heliocentric velocity at departure V_p is tangential to the orbit of the Earth. As a function of the eccentricity e , we have for an outbound orbit, following the presentation of Langevin (2005),

$$\begin{aligned} a &= 1/(1-e) \\ q &= (1+e)/(1-e) \\ \Delta M &= q \Delta V = (1+e)/(1-e) \Delta V \\ V_p &= V_e(1+e)^{0.5} = V_e + V_{\text{rel}} \end{aligned} \quad [13]$$

where V_e is the heliocentric velocity of the Earth and V_{rel} is the departure velocity

$$\begin{aligned} V_q &= V_p(1-e)/(1+e) \\ \Delta E_q &= V_q \Delta V = -V_p(1-e)/(1+e) \Delta V \end{aligned} \quad [14]$$

When returning to 1 AU (assuming a resonant orbit, so that the Earth is at the right place), the large reduction in angular momentum at aphelion results in a large reduction of the tangential velocity V_t with respect to its initial value (V_p)

$$V_t = V_p - (1+e)/(1-e) \Delta V$$

As the radial velocity is 0 at departure, if V_r is the radial velocity upon return, we have

$$\Delta E_p = 0.5(V_t^2 + V_r^2 - V_p^2)$$

Table 8 Equatorial radius and circular velocity of planets, corresponding to the orbital velocity on a grazing circular orbit

Mercury	Venus	Earth	Mars	Jupiter	Saturn	Uranus	Neptune
Radius (km)	2440	6050	6380	3400	71 000	60 000	25 500
$V_c (\text{km s}^{-1})$	3.01	7.33	7.91	3.55	42.1	25.1	15.1

For maneuvers which are small with respect to the heliocentric velocities, we can write

$$0.5(V_r^2 - V_p^2) = -V_p(1+e)/(1-e)\Delta V \quad [15]$$

There is an energy deficit at this stage, when comparing [15] and [14], which must be compensated by the radial component of the velocity V_r . Energy conservation leads to

$$\begin{aligned} \Delta E_p &= \Delta E_q \\ V_r^2 &= 2V_p\Delta V((1+e)/(1-e) - (1-e)/(1+e)) \\ V_r &= (8V_p\Delta V e/(1-e^2))^{0.5} \end{aligned} \quad [16]$$

As an example, for a 3-year resonant orbit with a departure in Jan at 6.84 km s^{-1} , the eccentricity is 0.527, the initial heliocentric velocity V_p is 37.14 km s^{-1} . The radial velocity at return is 2.16 km s^{-1} for a 22 m s^{-1} maneuver at aphelion and 4.60 km s^{-1} for a 100 m s^{-1} maneuver, close to the results of eqn [16]: 2.19 and 4.67 km s^{-1} , respectively. These small differences can be attributed to the slight eccentricity of the Earth orbit. When evaluating the relative velocity at Earth return, we have to combine this radial velocity with the tangential relative velocity, which is reduced by the same amount as the tangential heliocentric velocity. As long as V_r remains small with respect to V_{rel} , we have

$$dV_{\text{rel}}/\Delta V \sim 4V_p/V_{\text{rel}}e/(1-e^2) - (1+e)/(1-e) \quad [17]$$

The relative velocity at return for a 22 m s^{-1} maneuver is 7.104 km s^{-1} , a velocity increment of 260 m s^{-1} , while eqn [17] would result in a velocity increment of 277 m s^{-1} . Even with such a small maneuver, the radial velocity upon return V_r (2.16 km s^{-1}) is already not that small with respect to V_{rel} (6.84 km s^{-1}). The efficiency of the maneuver decreases as the radial velocity at the

Earth increases. Therefore, the ratio between the velocity increment and the ΔV (leverage factor) increases with the eccentricity of the orbit and it decreases with the velocity gain which is to be obtained.

A similar derivation for inbound orbits yields

$$V_p = V_e(1-e)^{0.5} = V_e - V_{\text{rel}} \quad [18]$$

$$dV_{\text{rel}}/\Delta V \sim 4V_p/V_{\text{rel}}e/(1-e^2) - (1-e)/(1+e) \quad [19]$$

For the same eccentricity e , an inbound DVGA is less favorable than an outbound DVGA, as V_p is now smaller than the heliocentric velocity of the Earth [18], contrarily to the outbound case [13]. The negative impact of the lower V_p/V_{rel} ratio is not compensated by the reduced value of the second term of [19] compared to that of [17].

Flying over the trailing side of the Earth (for outer solar system missions) or the leading side of the Earth (for inner solar system missions) rotates the relative velocity so that it is nearly aligned with the orbital velocity of the Earth after the encounter (Figure 12). This is most effective for reaching distant targets. The increment or decrement in heliocentric velocity over that of the Earth is the effective velocity, as it is equivalent to the departure velocity which would be provided by a direct launch in the direction of the heliocentric velocity of the Earth (or opposite to it for inner solar system missions). When the relative velocity is very large (9 km s^{-1} or more), a flyby at the lowest possible altitude ($\sim 250 \text{ km}$) is not able to fully rotate it to the optimum direction (see eqn [11]). The resulting effective velocity is then lower than the relative velocity, and the leverage factor decreases correspondingly. The most relevant examples are provided in Table 9. The launch ΔV is that which would have to be spent at the perigee of a

Table 9 Characteristics of relevant DVGA launch strategies

Resonance	Departure velocity (km s^{-1})	Launch ΔV (km s^{-1})	Eccentricity	Maneuver (km s^{-1})	Effective velocity (km s^{-1})	Total ΔV (km s^{-1})	Leverage factor	Additional cruise time (year(s))
1:1	-1.100	MGA	0.076	1.000	4.1	1.00	3.00	1
2:3	-5.342	1.286	0.329	0.428	7.55	1.72	5.16	2
3:4	-3.475	0.591	0.220	0.327	5.14	0.92	5.09	3
5:4	+2.020	0.236	0.141	0.246	3.41	0.48	5.64	5
4:3	+2.520	0.337	0.177	0.136	3.41	0.47	6.55	4
4:3	+2.586	0.352	0.181	0.485	5.17	0.84	5.33	4
3:2	+3.412	0.571	0.242	0.265	5.17	0.84	6.63	3
2:1	+5.171	1.214	0.377	0.224	6.98	1.43	8.07	2
2:1	+5.283	1.261	0.386	0.552	9.03	1.81	6.78	2
3:1	+6.983	2.095	0.523	0.195	9.03	2.28	10.49	3

Moon-crossing orbit so as to obtain the required departure velocity. For the cheapest orbit, a MGA can provide the required 1.1 km s^{-1} departure velocity (Farquhar, 1991), so that no additional ΔV is required.

The comparison between **Tables 8** and **1** demonstrates the potential of gravity-assist methods for the most demanding missions: a 9 km s^{-1} departure velocity required a velocity at perigee 3.29 km s^{-1} higher than that of a Moon-crossing orbit, which could not be achieved by a Fregat upper stage, and only marginally so with a Centaur upper stage. For such a high ΔV , a spacecraft engine burn around perigee (indirect launch strategy) would incur a staggering gravity loss. With a 2:1 DVGA, only 1.81 km s^{-1} are needed, with 1.26 km s^{-1} on the bound orbit so as to leave the Earth with a relative velocity of 5.28 km s^{-1} and 0.552 km s^{-1} as a deep space maneuver (for which there is no gravity loss). Using a 2:3 DVGA, only 1.71 km s^{-1} are needed to reach the 7.5 km s^{-1} negative velocity required for a direct Mercury flyby mission. With this approach, a spacecraft with a useful mass of about 1000 kg can be sent to Jupiter, a comet, or Mercury with the 2010 version of the Soyuz-Fregat (launched from Kourou). There is a penalty of 2 years in cruise time (hence a 5-year cruise to Jupiter), which is a small price to pay as such missions are now feasible with classical propulsion methods.

It would seem possible to further improve the mass budget by using a series of DVGA maneuvers. However, for departure velocities lower than 4 km s^{-1} , there is not much to be gained from a DVGA. the total ΔV budget required for a departure at 3.41 km s^{-1} starting from a Moon-crossing orbit is 0.48 km s^{-1} for a 5:4 DVGA and 0.47 km s^{-1} for a 4:3 DVGA, a 100 m s^{-1} improvement over the 0.57 km s^{-1} ΔV required for a direct departure from a bound orbit (the gravity loss is small at such low departure velocities). The 3% improvement of the mass budget is definitely not appealing considering the 4–5 years penalty in cruise time. Using a sequence of a 3:2 DVGA followed by a 2:1 DVGA for a Jupiter mission instead of a direct 2:1 DVGA is more attractive, as the ΔV reduction is now 0.42 km s^{-1} ($1.26\text{--}0.84 \text{ km s}^{-1}$), actually $\sim 600 \text{ m s}^{-1}$ as gravity losses become quite significant for departure velocities in the range of 4 km s^{-1} or more. The improvement of the mass budget is $\sim 20\%$, still a marginal proposition when considering the 3 years penalty in terms of cruise time. The 1:1 DVGA is specific as the departure velocity can be provided by a MGA. However, its leverage factor is low, and it is

interesting only for low-thrust missions such as BepiColombo (see Section 10.17.7).

With DVGA strategies, missions to Jupiter, comets, and Mercury would be within reach even if there was no other planet in the solar system. As we have seen in Section 10.17.3, Venus and Mars are relatively easy to reach. Tapping the gravity-assist potential of these two planets improves much the mass budget for more difficult targets, but there are additional constraints as now the Earth, Venus, or Mars, and the target body must be in the proper relationship.

Venus is particularly attractive as a stepping stone to the inner solar system as this planet is nearly as large as the Earth. Furthermore, it lies nearly halfway to Mercury both in terms of its semimajor axis (0.723 AU, when Mercury is at 0.387 AU) and its orbit plane (inclination of 3.39° , to be compared with the 7° inclination of Mercury's orbit). A close swingby of Venus changes the velocity by as much as 7 km s^{-1} , which is more than that can be provided by a single rocket stage. All missions to Mercury therefore include at least one swingby of Venus, with departure velocities from the Earth in the $3.5\text{--}4 \text{ km s}^{-1}$ range, much lower than the 7.5 km s^{-1} which would be required for a direct transfer. Indeed, the Mariner 10 flyby mission to Mercury was the very first mission to use a planetary gravity-assist, with a launch in Nov 1973, a flyby of Venus in Feb 1974, and three flybys of Mercury in 1974 and 1975 (Shirley, 2003). For rendezvous missions, two Venus swingbys are required (Yen, 1989): the first swingby sets the spacecraft into a resonant orbit. The shortest solutions involve the 1:1 resonance (period of 0.615 years), the cheapest solutions use the 3:4 resonance (period of 0.461 years, coming back to Venus after 1.84 years). A second swingby then sends the spacecraft to Mercury which is reached after 1.5–3.5 full orbits around the Sun. (Mercury is usually not at the best position for a transfer, which requires inserting ‘phasing’ orbits.) The arrival velocity is $\sim 5.8 \text{ km s}^{-1}$ in the best cases. This is still far too high for an orbiter mission: the corresponding capture maneuver would be in the 3 km s^{-1} range, and could not be accommodated in the short time available close to this small planetary body (2450 km in radius). Fortunately, a sequence of gravity assists can also be used for reducing the arrival velocity, with the same leverage factors as for pumping up the departure velocity. This strategy is particularly effective for Mercury, which has a short orbital period (0.24 years). Mercury mission strategies developed

by Yen (1989) involve a series of three DVGAs (2:1, 4:3, 6:5) with a total cost of 500 m s^{-1} and a duration of 12 Mercury periods (slightly less than 3 years). This reduces the relative velocity at arrival from 5.8 km s^{-1} down to 2.2 km s^{-1} , at which stage an insertion maneuver to a highly elliptical orbit requires only 500 m s^{-1} . Mercury has no atmosphere. Therefore, one has to rely on the spacecraft engine to further reduce the orbital period. The total ΔV cost of a Mercury orbiter on a 10 h elliptical orbit is now in a manageable range ($1.5\text{--}2 \text{ km s}^{-1}$) which brings more than 50% of the launch mass into orbit, but cruise durations now range from 5.6 years up to more than 9 years depending on the Venus window and the Venus–Mercury relative positions. Such a strategy is being implemented by the MESSENGER mission of NASA (Santo *et al.*, 2001; McAdams *et al.*, 2006), launched in 2005 (Figure 4).

Mars is not a very useful stepping stone to the outer solar system, as its mass is much smaller than that of Venus (the maximum ΔV available from a Mars gravity-assist is $\sim 3 \text{ km s}^{-1}$). Furthermore, the orbit of Mars is much closer to that of the Earth than

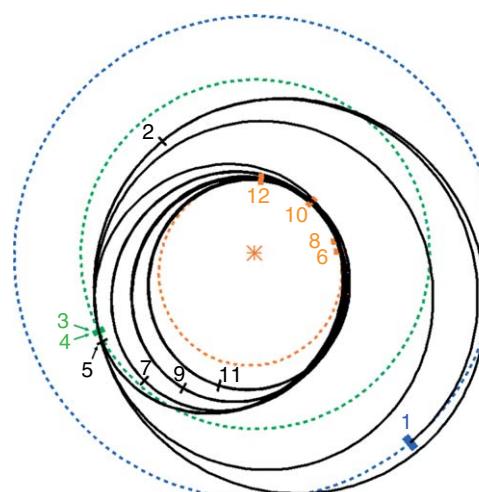


Figure 4 MESSENGER mission to Mercury, launched in Aug 2005, with an insertion in Mercury orbit in Mar 2011. After the launch at 4.05 km s^{-1} (1) and a deep space maneuver (2) which increases the relative velocity to Venus, two Venus swingbys are implemented (3, 4). A small maneuver (5) is needed to reach Mercury (6). After the first flyby of Mercury (6) with a relative velocity of 5.8 km s^{-1} , a 3:2 DVGA (6-7-8), a 4:3 DVGA (8-9-10), and a 6:5 DVGA (10-11-12) are implemented. These three maneuvers lower the relative velocity to 2.2 km s^{-1} at a cost of only 500 m s^{-1} in ΔV (but more than 3 years in cruise time). The insertion maneuver into MESSENGER's elliptical orbit is now at an acceptable level (870 m s^{-1}).

to that of Jupiter. Conversely, Jupiter has such a large circular velocity (42 km s^{-1}) that it can sling the spacecraft towards any target in the outer solar system if the relative position is favorable. The Voyager 1 mission, launched in 1977, took advantage of a perfect configuration to visit Jupiter, Saturn, Uranus, and Neptune from 1980 to 1989, using each planet as a stepping stone to the next target. This favorable configuration repeats only once in 170 years. Given the relative periods of Jupiter (11.9 years) and Saturn (29.6 years), a favorable configuration between Jupiter and Saturn repeats every 19.8 years. The next visit to Saturn, the Cassini/Huygens mission, was launched in 1997, 20 years after Voyager, and the next opportunity is in 2015–17. Such a mission to Saturn using first a 2:1 DVGA to reach Jupiter is presented in Figure 5. It could be considered for a post-Huygens mission to Titan in ‘Cosmic Vision’, the 2015–20 program of ESA.

The best use of the gravity potential of Venus and Mars for outer solar system missions is to replace the deep space maneuver by a gravity-assist in an Earth DVGA. One can then reach large departure velocities from the Earth without expending any onboard ΔV , which improves much the mass budget (see, e.g., Longuski and Williams, 1991; Petropoulos *et al.*, 2000). A Mars flyby is most effective when providing the deep space maneuver in a 3:2 Earth DVGA scenario. The cheapest opportunity, with a departure velocity of $\sim 3.2 \text{ km s}^{-1}$, occurs every 2.1 years, and the Earth return is 2.8 years later at a relative velocity

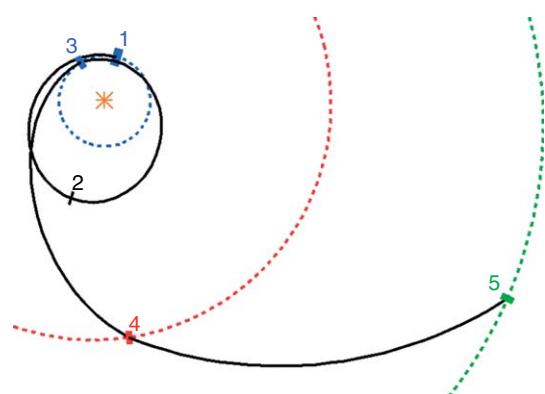


Figure 5 2:1 DVGA mission to Saturn to be launched in late 2015 (1) with a departure velocity of 5.2 km s^{-1} . A deep space maneuver (2) 565 m increases the relative velocity with respect to the Earth to 9.25 km s^{-1} . An Earth gravity-assist (3) sends the spacecraft to Jupiter. A Jupiter gravity-assist (4) sends the spacecraft to Saturn (5) which is reached in late 2023 at a low relative velocity of 5.1 km s^{-1} .

of 9 km s^{-1} (Hechler, 1997). A 2 years intermediate orbit may be needed so as to adjust the inclination. The cruise time is then increased by 5 years. A version of this strategy was used for the 2004 launch of Rosetta by Ariane 5. An initial 1:1 DVGA was added so as to optimize the launch declination while increasing the relative velocity from 3.4 to 3.85 km s^{-1} (Figure 6). With this strategy, the total ΔV is only 1.68 km s^{-1} , and the useful mass in rendezvous is 1400 kg , which makes it possible to implement a large orbiter and a lander on the comet nucleus. The drawback is the cruise time of 10.5 years. This tradeoff between mass budget and cruise time, already encountered for Mercury orbiter missions, is inherent to gravity-assist strategies: given enough time, very little is impossible.

A Venus gravity-assist can provide for free a very large ΔV (up to 7 km s^{-1}). Therefore, a Venus flyby is already quite effective when used as the maneuver in a 1:1 Earth DVGA. Indeed, every 1.61 years, it is

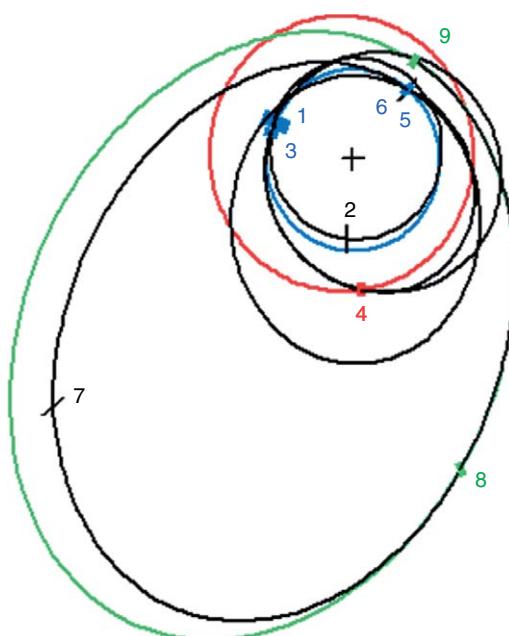


Figure 6 Trajectory of the Rosetta spacecraft to comet Churyumov-Gerasimenko (green orbit). After a launch by Ariane 5 in late Feb 2004 with a departure velocity of 3.4 km s^{-1} (1); a 1:1 DVGA maneuver (2) increased the relative velocity at the Earth to 3.85 km s^{-1} (3); a 3:2 Earth–Mars–Earth gravity-assist (3-4-5) increased the velocity of the Earth encounter to 9.35 km s^{-1} (5); a sequence of two Earth flybys (5-6) is required to tilt the orbit so as to match the inclination of the comet. Two deep space maneuvers (7-8) totaling 1.45 km s^{-1} implement the rendezvous at 3.7 AU from the Earth in Sep 2014 (8). The comet is followed until perihelion (9) in Oct 2015 in the nominal mission.

possible to depart to Venus with a velocity of $3.2\text{--}3.7 \text{ km s}^{-1}$, swingby Venus at a very safe distance, then return to the Earth with a relative velocity of $7.5\text{--}8.3 \text{ km s}^{-1}$ after 1 year. The period of the orbit is very close to 1 year. Hence, this relative velocity is nearly perpendicular to the orbital velocity of the Earth, when the maximum rotation angle from an Earth flyby at this velocity as derived by eqn [11] is only 50° . Therefore, a first Earth gravity-assist is used to transfer the spacecraft to an orbit with a 2 years period, during which a small maneuver can be implemented so as to further increase the relative velocity (2:1 DVGA). The second Earth gravity-assist sets the spacecraft on its way to Jupiter or the target comet. The cruise time is 3 years longer than for a direct transfer, but the low departure velocity provides large mass budgets (nearly 4 tons approaching Jupiter with Atlas V-421 + Centaur). This strategy was first implemented for sending the Galileo spacecraft into orbit around Jupiter (D'Amario *et al.*, 1992), as it was much larger (1.6 tons dry mass, 2.55 tons at departure) than Pioneer or Voyager. An example of such a scenario is a mission to the system of Jupiter to be launched in late 2016 which could be considered as a candidate for the 2015–20 program of ESA (Figure 7), as the mass budget is very favorable. A different type of Venus–Earth gravity-assist strategy followed by a Jupiter swingby was used for the Cassini/Huygens mission

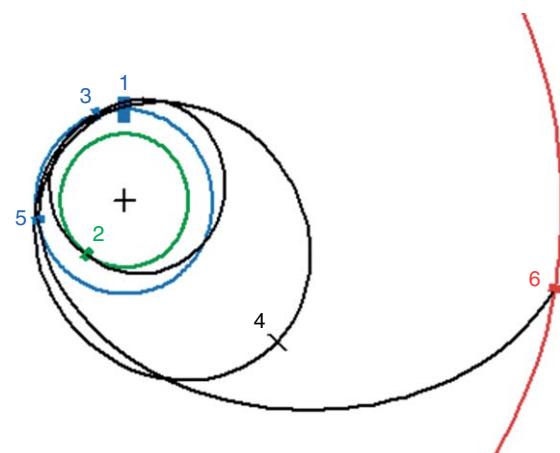


Figure 7 Venus–Earth–Earth gravity-assist mission to Jupiter to be launched in late 2016. After a launch at 3.2 km s^{-1} (1), a swingby of Venus (2) increases the relative velocity to 7.7 km s^{-1} when returning to the Earth (3). A 2:1 DVGA maneuver of 235 m s^{-1} (4) increases the relative velocity to 9.2 km s^{-1} at the next Earth flyby (5). Jupiter is reached after 5.4 years (6). The Galileo spacecraft followed a very similar route to Jupiter from 1989 to 1995.

to Saturn and Titan launched in 1997 (Peralta and Flanagan, 1995). The cheap trip to Jupiter of **Figure 7** is not optimal for a mission to Saturn as Jupiter is too far ahead on its orbit when compared to the favorable configuration of **Figure 5**. It is indeed possible to reach Saturn for free with a Jupiter swingby, but the arrival would be in 2028 even with a ΔV expenditure of 630 m s^{-1} instead of 235 m s^{-1} to Jupiter.

Gravity-assists provide very effective means of exploring the environment of a planet if one or more large natural satellites are present. A remarkable example is the 4 years tour of the Saturn system by Cassini (Wolf and Smith, 1995), with its 50 flybys of Titan, the largest satellite of Saturn. All other major satellites are visited, the rings are observed at close range, and high latitudes in the magnetosphere are explored. The Jupiter system provides outstanding opportunities (Longman and Schneide, 1970) with its four Galilean satellites (Io, Europa, Ganymede, and Callisto), all with a large gravity-assist potential as escape velocities are larger than 2 km s^{-1} (**Table 5**), hence velocity changes of $\sim 1 \text{ km s}^{-1}$ can be achieved if the relative velocity is not too large. This was already demonstrated by the Galileo mission of NASA (D'Amario *et al.*, 1992). A complex sequence of DVGA and gravity-assist strategies between Galilean satellites can be implemented in relatively short times, as the period of Ganymede (third from Jupiter) is only 7 days instead of 365 days for the Earth (third from the Sun).

A mass-effective scenario can be designed for inserting a spacecraft into orbit around one of the Galilean satellites (e.g., Europa). A sequence of Io gravity-assists reduces the period of the orbit after insertion. Two Io–Callisto–Io gravity-assist sequences lower the velocity relative to Io from 14 to less than 6 km s^{-1} . It is then possible to implement an Io–Ganymede–Ganymede–Europa sequence followed by a 3:2, 7:6 DVGA sequence from Europa to Europa (**Figure 8**) which lowers the velocity relative to Europa down to 660 m s^{-1} . A ΔV of 675 m s^{-1} is then required for reaching a circular polar orbit around Europa. This strategy, with two Callisto flybys, two Ganymede flybys, two Europa flybys, and 10 Io flybys is still relatively simple compared to the Cassini tour (60 flybys during the nominal mission alone). It requires ~ 1 year after orbit insertion. If one implements this strategy after reaching Jupiter as described in **Figure 7**, the required total onboard ΔV including the insertion maneuver and 200 m s^{-1} for navigation is in the range of 2 km s^{-1} . With the version of Ariane 5 used for Rosetta (3500 kg at

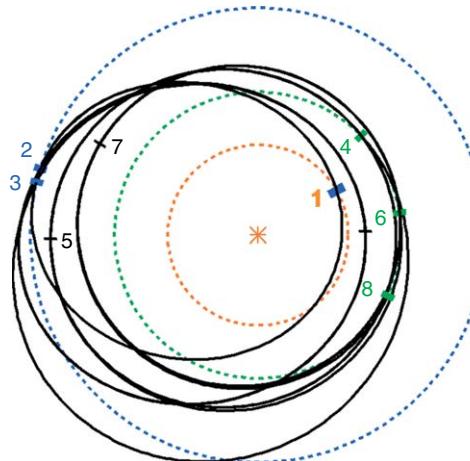


Figure 8 End of mission example for a Europa orbiter mission exploiting the gravity-assist potential of Galilean satellites Io (orange), Europa (green), and Ganymede (blue). After a series of Io swingbys which reduce the orbital period, and a Io–Callisto–Callisto–Io gravity-assist sequence which reduces the relative velocity to Io (1), a double swingby of Ganymede (2-3) raises the pericenter close to the orbit of Europa (4). A 3:2 DVGA (4-5-6) then a 6:5 DVGA (6-7-8) reduces the relative velocity to Europa from 1.91 km s^{-1} to 0.66 km s^{-1} . At this stage, a cheap insertion in a low circular polar orbit by the onboard propulsion system becomes possible. The orbit insertion around Europa is performed only 60 days after the departure from Io, and the total cost of maneuvers is small (225 m s^{-1}).

3.2 km s^{-1}), the useful mass in orbit around Europa could reach 1500 kg, including, for example, 900 kg of shielding against the extreme radiation environment of the inner Jovian system.

10.17.8 Advanced Propulsion Systems: Solar Sail, Ion Propulsion

Gravity-assist techniques presented in Section 10.17.7 have made possible a wide range of missions which would have been out of reach of a direct launch strategy. It is therefore interesting to present the arguments for new propulsion techniques. It should first be noted that several missions of interest remain nearly impossible or require very long cruise times:

1. rendezvous with a small body (asteroid or comet) if the perihelion is not close to 1 AU, or if the inclination exceeds 15° (the rendezvous maneuver exceeds the 3 km s^{-1} limit); this is the case for all main-belt asteroids, and all but a few comets;

2. missions to Saturn as the largest mass budgets are available only once every 20 years (when Jupiter can be used as a stepping stone) or with exceptionally favorable Venus–Earth opportunities; and
3. rendezvous missions to Uranus, Neptune, and KBOs (including Pluto).

An additional problem is the exceedingly long cruise time, with an impact on the risk/reliability tradeoff (hence cost). The relevance of missions designed 20 years before the science return is obtained is also an issue. Such missions cannot benefit from the latest advances in instrument design. They cannot adjust to new discoveries from other solar system exploration missions. Furthermore, there is an important programmatic risk, as other space agencies may try to steal the fire out of the mission with a smaller, faster mission. As an example, NASA programmed, then cancelled a comet lander mission which would have arrived 5 years earlier than the lander of Rosetta. The complexity of the Rosetta mission strategy (**Figure 6**) demonstrates that gravity-assist techniques have reached their limit with classical propulsion systems.

The only solution for improving this situation is a major increase in onboard ΔV capability, which requires increasing the ejection velocity V_{ej} . There is however a catch. The thrust T is

$$T = (\text{dm}/\text{dt})V_{\text{ej}} \quad [20]$$

dm/dt is the mass flux in kg s^{-1} . The corresponding power W is

$$W = 0.5(\text{dm}/\text{dt})V_{\text{ej}}^2 \quad [21]$$

hence

$$W = 0.5V_{\text{ej}}T \quad [22]$$

For a given thrust, the required power is therefore proportional to the ejection velocity. Large ejection velocities are appealing as they reduce the fuel mass for a given ΔV , but one must be aware of the larger corresponding energy requirements.

10.17.8.1 Solar Sail Propulsion

The ‘photon drive’ (e.g., a laser) is popular in science fiction novels, and it provides the highest possible ejection velocity (c , the speed of light). At relativistic speeds, the relationship between energy and thrust is less favorable than that given in eqn [22], as we have

$$W = cT \quad [23]$$

From eqn [23], an effective power of 300 MW would be required for a modest 1 N thrust (420 MW onboard power if the conversion efficiency is $\sim 70\%$). Unless one considers direct mass-energy conversion (yet another science fiction favorite), the only realistic propulsion method using photons consists in implementing a solar sail, which reflects solar photons. This looks very appealing, in particular for very demanding missions such as sample returns ([Hughes et al., 2006](#)), as the photons are provided for free by the Sun, hence there is no mass expenditure whatever the total ΔV . However, the available power at 1 AU is 1400 W m^{-2} , hence a $200\,000 \text{ m}^2$ sail (nearly $500 \text{ m} \times 500 \text{ m}$) is required for a 1 N thrust level. A maximum of 2 g m^{-2} for the sail and structure is required so as to keep the mass of the sail below 1000 kg. Furthermore, reflective properties must be maintained against high radiation levels for the full cruise phase ([Edwards et al., 2004](#)). Due to these difficulties, no solar sail mission has yet been flown. Several test missions have been considered by NASA and ESA, but the first test of solar sail propulsion is yet to be approved.

10.17.8.2 Ion Propulsion

Accelerating ions instead of molecules (ion propulsion) provides ejection velocities in the $15\text{--}100 \text{ km s}^{-1}$ range which are much higher than that provided by chemical thrusters (3.1 km s^{-1} for onboard systems). The best fuel is a heavy noble gas, such as xenon. There are two methods for accelerating ions: the expansion of a hot plasma generates ejection velocities in the lower part of the range ($15\text{--}20 \text{ km s}^{-1}$). With a high-voltage grid, it is possible to accelerate individual ions to velocities in the $30\text{--}100 \text{ km s}^{-1}$ range. Higher ejection velocities require less fuel for a given ΔV . However, the required installed power increases with ejection velocity, which has an impact on the mass of the power subsystem. Electric energy can be obtained in space either from solar panels (SEP) or from a small nuclear reactor (NEP).

At 1 AU, a solar panel provides $\sim 280 \text{ W m}^{-2}$ (with a 20% efficiency, which is already quite good). This corresponds to 26 mN with a 15 km s^{-1} ejection velocity, 8 mN with a 50 km s^{-1} ejection velocity, assuming an efficiency of 70% for the ion acceleration process. An SEP propulsion system is therefore limited to relatively small thrusts (in the range of 100–200 mN at 1 AU), otherwise the solar panels become very large, which has an impact in terms of

mass and cost. A 200 mN thrust seems very small, but if it is sustained over 1 year, it can provide 3.15 km s^{-1} of ΔV to a 2000 kg spacecraft. SEP missions are therefore characterized by very long thrust arcs.

SEP systems have been successfully tested by NASA (*DS-1*, which encountered comet Borelly), ESA (Smart-1, in orbit around the Moon in 2005 and 2006; [Racca et al., 2002](#)) and Japan (Hayabusa, which performed a rendezvous with asteroid Itokawa in 2005). Two major missions are planned in the next 10 years: Dawn (NASA), launched in 2007, will rendezvous with two of the largest main-belt asteroids, Vesta in 2011, then Ceres in 2015; BepiColombo (ESA/Japan) is a dual orbiter mission to Mercury to be launched in 2013.

Even if onboard ΔV is more plentiful than with chemical missions, gravity-assist strategies can also be quite interesting for SEP missions, as demonstrated by the evolution of the mission scenario for BepiColombo, which initially included two orbiters and a lander. The very first studies considered a direct spiral from the orbit of the Earth down to the orbit of Mercury of a large composite spacecraft with all three science modules. Even with an Ariane 5 launch, this required more than 16 km s^{-1} of ΔV with relatively high thrust levels (400 mN), and a 2.5-year cruise time. These requirements exceeded the lifetime of thrusters, so that full spares were needed with a drastic impact on the cost and mass budgets, and this option was abandoned. Inserting a double Venus swingby reduced the ΔV by $\sim 7 \text{ km s}^{-1}$, and a final 180° transfer from Mercury to Mercury, followed by a 1:1 DVGA saved an additional 1.5 km s^{-1} ([Langevin, 2000](#)), so that the ΔV budget went down to a manageable level of $\sim 7.5 \text{ km s}^{-1}$ ([Figure 9](#)).

As usual in the development of a project, the mass requirements of subsystems rised significantly (in particular for a Mercury lander, see Section 10.17.6) and cost became an issue. Two Soyuz–Fregat launches were considered, one for the lander and one for the two orbiters. The largest mass is then obtained by launching on a Moon-crossing orbit ($\sim 2200 \text{ kg}$). A gravity-assist by the Moon provides for free a departure velocity of $1\text{--}1.4 \text{ km s}^{-1}$. A 1:1 DVGA then builds up a large relative velocity when returning to the Earth at a moderate cost in terms of onboard ΔV (typically 1 km s^{-1} for a 3.5 km s^{-1} relative velocity). With a 45 km s^{-1} ion ejection velocity, the remaining mass is 2150 kg, 400 kg more than the 1750 kg which would be available with a direct launch. As usual, mass costs time, in this case 1.3 years for the round trip back to the Earth.

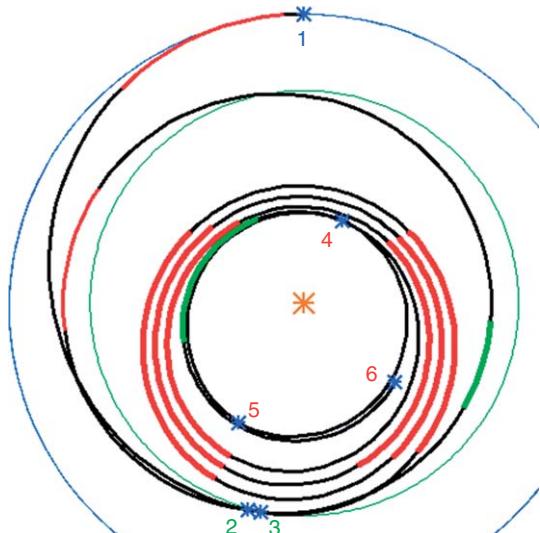


Figure 9 Direct launch option by an Ariane 5 for the BepiColombo mission to Mercury in late 2009. The departure velocity of 2.5 km s^{-1} is increased by early thrusting with the ion engine, so as to reach Venus with a relative velocity of 4.5 km s^{-1} (2). A 1:1 Venus DVGA increases the relative velocity to 5.7 km s^{-1} (3). After a three orbit spiral, Mercury is reached at a relative velocity of 2.7 km s^{-1} (4). A 180° transfer (4-5), then a 1:1 Mercury DVGA (5-6) reduces the relative velocity to less than 400 m s^{-1} . Red arcs correspond to thrust against the orbital velocity, green arcs correspond to thrust along the orbital velocity. The cruise time is 2.6 years.

The cost restrictions on cornerstone missions led in 2003 to the cancellation of the lander, which required a full Soyuz launch with a dedicated cruise module and an ion propulsion module (see Section 10.17.6). The new configuration of the mission consisted of two orbiters launched by a single Soyuz–Fregat. The maximum thrust of the SEP propulsion system was reduced to 240 mN, so as to save mass and cost, but this had an adverse effect on the ΔV budget. In parallel, subsystem mass (in particular that of the SEP stage) continued to grow, which led to a mass crisis for the Soyuz–Fregat launch of the two orbiter composite. New scenarios had to be considered so as to further reduce ΔV , hence fuel mass. One of these options, with a total ΔV of 6.3 km s^{-1} is to be launched in Nov 2013. The other, with a total ΔV of 4.15 km s^{-1} is to be launched in May 2014 ([Figure 10](#)), with two additional Mercury–Mercury DVGA sequences (3:2 and 5:4). There are now similarities with a ballistic mission such as MESSENGER ([Figure 4](#)). Thanks to the low fuel consumption of electric propulsion at the selected

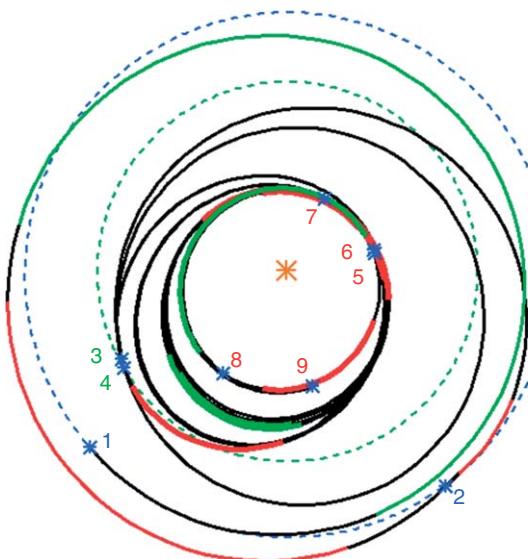


Figure 10 Cheap transfer to Mercury with a Soyuz–Fregat launch in May 2014. Very little braking thrust is now applied on the way from Venus (4) to Mercury (5). The relative velocity, which is 5.6 km s^{-1} at that stage, is reduced to 2.4 km s^{-1} by a 3:2 DVGA (5-6) then a 5:4 DVGA (6-7). The end-of-mission strategy (7-8-9) is adjusted to the lower thrust-to-mass ratio.

ejection velocity (45 km s^{-1}), more than 1900 kg of the 2200 kg launched in a Moon-crossing orbit can be delivered at a very low relative velocity to Mercury, but the cruise time is now close to 6 years.

The evolution of BepiColombo illustrates the remarkable flexibility of electric propulsion: contrarily to chemical propulsion missions, the large onboard ΔV capability makes it possible to adjust to major changes in mass budgets or launch dates. The combination of gravity-assists with low thrust is extremely effective, as the initial ΔV budget of 16 km s^{-1} could be reduced to $\sim 4.15 \text{ km s}^{-1}$, albeit with a significant increase in cruise time (from 2.5 to 6 years). Additional information on SEP optimization methods can be obtained from [Chesta et al. \(2006\)](#), while the navigation aspects of BepiColombo are addressed by [Yarnoz et al. \(2006\)](#).

SEP is only effective relatively close to the Sun. The Dawn mission, which implements a rendezvous with Vesta, at 2.4 AU, then Ceres, at 2.8 AU, already requires very large solar panels. Therefore, SEP is totally inappropriate for outer solar system missions. Nuclear reactors are the only power source which can provide several 10 kW at large distances from the Sun, so as to power ion thrusters (NEP). The fissile material must exceed a critical mass. Therefore, the

smallest nuclear reactors provide several 100 kW, with a mass of several tons. The scale of NEP missions is daunting: Jupiter Icy Moons Orbiter (JIMO), the mission to Galilean satellites which was studied, then cancelled by NASA, consisted of a 8 tons spacecraft, with 5 tons for the reactor and shield, and a thrust of 8 N. Beyond Jupiter, the acceleration of 1 mm s^{-2} far exceeds the gravitational pull of the Sun (0.23 mm s^{-2} at 5 AU). Therefore, trajectories are non-Keplerian, which makes it possible to reach Saturn, Uranus, or Neptune in a fraction of the time which would be required on a Hohmann transfer orbit (8 years for Saturn, 29 years for Neptune). Such fast trajectories require ΔV budgets $> 20 \text{ km s}^{-1}$, as one has to accelerate, then brake at arrival. Fortunately, ejection velocities can be very high ($\sim 100 \text{ km s}^{-1}$) as there is a lot of available power, hence 20 km s^{-1} of ΔV only uses up $\sim 20\%$ of the mass as fuel. The demise of JIMO, initially planned for the 2012 time frame, is a severe setback for NEP, which is undoubtedly an enabling technology for a systematic exploration of the outer solar system.

10.17.9 The Specific Challenges of Sample Return Missions

Implementing a sample return mission is a daunting task, as one can consider it as a rendezvous mission followed by its mirror image. This view is to a large extent correct. However, there are some mitigating circumstances. First of all, the Earth has a dense atmosphere, which is quite capable of braking sample return modules at large velocities. It is ‘only’ required to set the spacecraft (or the sample return capsule, provided that it can survive alone) on a trajectory which intersects that of the Earth. In terms of ΔV budget, one gets rid of the huge cost of leaving the sphere of influence of the Earth. However, one has to launch from the target body.

Asteroids and comets represent relatively easy targets for a landing and sample collection, as escape velocities are either low or negligible. A specific case is constituted by the Stardust mission, which is expected to bring back samples from comet Wild-2 in early 2006. After a launch in early 1999, a 2-year DVGA maneuver increased the relative velocity so that an Earth gravity-assist in Jan 2001 set Stardust on its way to Wild-2 on an orbit with a period of nearly 2.5 years. The samples were collected at a high velocity (6.1 km s^{-1}) during a flyby in early 2004, and

a small maneuver was then required to encounter the Earth in Jan 2006, 5 years after the Earth gravity-assist, after two full orbits. Due to the high impact velocity, only very small grains can be captured relatively unaltered. For a comet nucleus sample return (once the baseline of Rosetta which was then considered in collaboration with NASA), two large maneuvers are needed so as to rendezvous with the comet, then leave it towards the Earth. **Figure 4** illustrates the specific difficulty of implementing a comet nucleus sample return with a chemical propulsion mission: cheap rendezvous arrive after aphelion, and by symmetry cheap returns depart before aphelion. One would need to wait nearly a full orbit around the comet. Hence the duration would be that of a rendezvous mission plus 6 years (typical period for a comet) plus 3 years for the return, for a total of 15–20 years. Much shorter sample return missions can be achieved with ion propulsion, as the rendezvous then occurs after perihelion, hence the return trip can be initiated almost immediately (e.g., Kawaguchi, 2002). An example is given in **Figure 11** for comet Wild-2. It constitutes a first illustration of the mirror image character of the return trip. The 1:1 DVGA loop after launch can be skipped for the

return leg as there is no need to reduce the relative velocity to the Earth, which will be cheaply dissipated in the atmosphere of our planet.

The very first SEP sample return mission should have been the Japanese Hayabusa missions. It achieved its rendezvous with asteroid Itokawa in Sep 2005, but it did not succeed in collecting a sample and bringing it back to Earth and the Japanese space agency (JAXA) is now considering a reflight. As Itokawa is an Earth-crossing asteroid, the ΔV budget was relatively low. A sample return from main-belt asteroid such as Vesta would have similar requirements ($\Delta V \sim 10 \text{ km s}^{-1}$) and the same mirror spiral trajectory as a sample return from a cometary nucleus.

Among large bodies, the Moon is by far the easiest target for a sample return mission. It has a relatively low escape velocity and surface gravity (see **Table 5**) which more than offset the lack of an atmosphere. Furthermore, the trip to the Moon and back is very cheap: a Moon-crossing orbit is obviously adequate. The approach (or departure) velocity is 0.8 km s^{-1} , so that the ΔV required to reach or leave a low circular lunar orbit from a transfer orbit is only 0.8 km s^{-1} (by coincidence). **Table 7** shows that the mass ratio between the surface and escape is 3.5 (4 with an arrival or departure velocity of 0.8 km s^{-1}). Therefore, a 50 kg return capsule requires only about 800 kg approaching the Moon, with a two-stage approach (one for each leg). Indeed, the Soviet missions Luna 16, Luna 20, and Luna 24 of the 1970s are until now the only successful automatic sample return missions from a large body (including a 2 m-deep core), at a very small fraction of the cost of the Apollo program.

The next easiest targets for a sample return mission are expected to be the two planets which are easiest to reach for a rendezvous, Mars and Venus. In both cases, it is possible to use aerobraking at arrival, which improves the mass budget. However, Mars and even more so Venus are large planets, with escape velocities of 5.02 and 10.36 km s^{-1} , respectively.

The escape velocity of Mars is even larger than that of Mercury (4.26 km s^{-1}). As we have seen in Section 10.17.6, such levels of ΔV are almost out of reach of a single stage. Tradeoffs for such a mission are presented in Donahue *et al.* (2006). The circular velocity (3.553 km s^{-1} , at the surface) is more manageable, as it is closer to the exhaust velocity of bipropellant propulsion systems. The logical scenario for a Mars sample return mission is therefore that which was considered for a joint mission between NASA and several European countries (France, Germany, Finland, etc.).

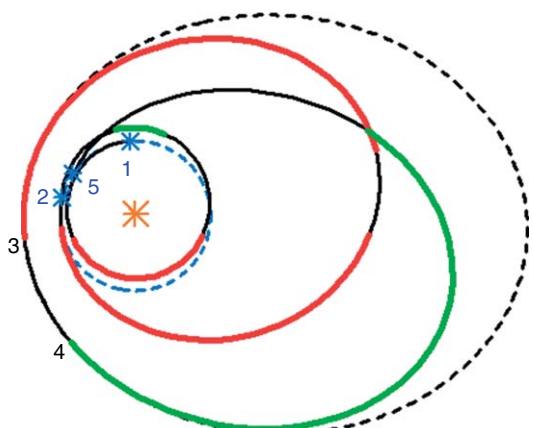


Figure 11 SEP comet sample return mission to Wild-2. A lunar swingby in Dec 2016 provides a departure velocity of 1.4 km s^{-1} (1). A two arcs 1:1 DVGA increases the relative velocity to 5.9 km s^{-1} (2). Rendezvous (3) is to be achieved in Sep 2021 by accelerating nearly continuously for 2.5 years (red arcs). After 3 months near the comet, the SEP propulsion system is reactivated, braking during 3 years (long green arc) so as to achieve a return to the Earth in Feb 2026(5). The total ΔV is an acceptable 10.14 km s^{-1} , the round trip duration is 9.14 years. Eighty percent of the mass is available for the SEP stage and the useful mass, so that a high-capability launcher such as Ariane 5 may not be required for this mission.

1. A large orbiter spacecraft is launched towards Mars, where it implements an aerocapture maneuver into a low circular orbit so as to maximize the mass in orbit. It carries the return capsule.
2. A large direct entry module is launched independently. It carries a small surface to orbit module. Once on the surface, samples are collected, then transferred to the surface to entry module, which sends a small sample cannister in orbit.
3. The orbiter collects the sample cannister, which is transferred to the sample return module.
4. When the return window opens, the sample return module is launched towards the Earth.
5. The sample cannister is recovered after a direct entry into the Earth's atmosphere. In the case of Mars, an aerocapture into low Earth orbit was considered so as to implement a quarantine in the International Space Station, but it is now unclear whether this facility will still be in operation when the samples come back, in 2018–20 at the earliest.

Figure 12 provides an example of a Mars sample return mission to be launched in late 2013. The symmetry of the mission design is immediately apparent. The ΔV budget for the return is in this case very low (2.03 km s^{-1} from a circular orbit to a departure velocity of 2.5 km s^{-1} toward the Earth). A

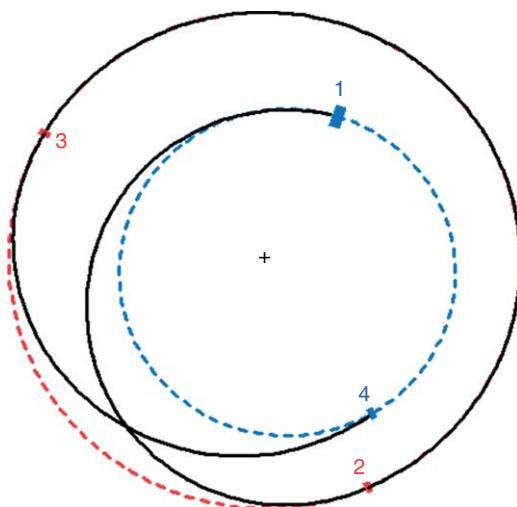


Figure 12 Mars sample return mission to be launched in Dec 2013 (departure velocity: 3.1 km s^{-1}), which could be considered as the second mission of ESA's exploration program. A type II transfer reaches Mars in Sep 2014. One must wait 1.15 years in orbit, so as to reach a very favorable type I return window, with a departure in Nov 2015 at 2.5 km s^{-1} and sample delivery in Aug 2016, with a moderate approach velocity (5.4 km s^{-1}). The round trip takes 2.7 years.

mass budget of $\sim 200 \text{ kg}$ is probably achievable for the sample return module. Arrival at Mars (2) is during southern summer storm season, which is not a problem for the orbiter, but may constitute a risk for the lander. A launch in Sep 2013 with one full additional orbit around the Sun arrives earlier in the seasonal cycle, but the Mars operational sequence would need to be completed in 2 months so as to catch the end of the return window in early 2016, and even then the departure velocity would be 3.5 km s^{-1} , with a higher ΔV 2.5 km s^{-1} (plus gravity losses) for the sample return module.

Venus is a relatively easy case for the round trip issue. **Figure 13** gives an example of such a mission to be launched in late 2016, which takes nearly 2 years. The problem is that Venus is almost as large as the Earth (escape velocity of 10.35 km s^{-1} instead of 11.78 km s^{-1}). The hot and very thick atmosphere (90 bars at the surface, 450°C) adds to the difficulty. At arrival, one can use aerocapture so as to retain a large fraction of the mass budget in low circular orbit or on the ground. At departure, the sample return module will need a ΔV of 3.36 km s^{-1} (plus gravity losses) from a low circular orbit, which requires by itself a large propulsion system. Recovering anything

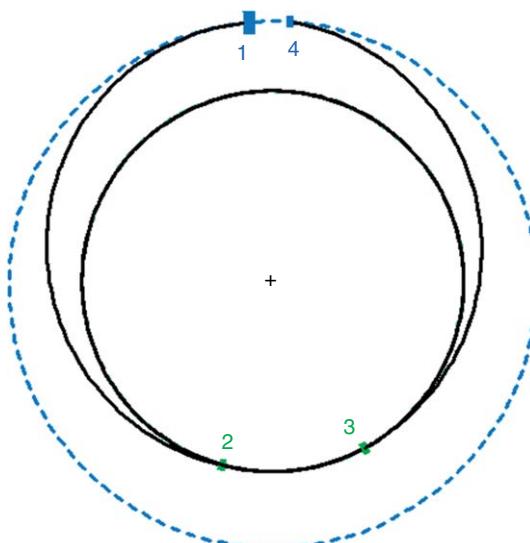


Figure 13 Venus sample return mission to be launched in Dec 2016 (departure velocity: 2.7 km s^{-1}). A type I transfer reaches Venus in Apr 2017. One must wait more than two full periods of Venus (1.2 years from 2 to 3) so as to reach a favorable type I return window, with a departure in Aug 2018 at 2.93 km s^{-1} and a sample delivery in Dec 2018, with a low approach velocity (4.3 km s^{-1}). The round trip takes slightly less than 2 years. In this example, the symmetry between the departure and return legs is particularly clear.

from the surface is a major issue: this is energetically similar to a launch in Earth's orbit, with mass ratios higher than 20, but the launcher has to start under the equivalent of 900 m of water (400 m at the top of Maxwell Montes, the highest mountain of Venus). Until now the record of survival at the surface, held by a Soviet Venera lander, stands at 2 h. It is therefore clear that enabling high-temperature technologies need to be validated by successful long-duration *in situ* missions (see, e.g., Landis, 2006) before addressing the daunting mass-budget challenges of a Venus sample return mission from the surface itself. Recovering a sample from the upper atmosphere is a much less challenging task (Sweetser *et al.*, 2003).

When considering the major technical issues which had to be solved for an ambitious orbiter mission to Mercury and the very large mass budget required to land any useful mass on the surface, a sample return from Mercury appears very far-fetched indeed, unless solar sail missions really come of age (McInnes, 2003). The same applies to large natural satellites of giant planets, where major steps in terms of orbiter mission and *in situ* studies need to be achieved before a sample return can be considered. Therefore, the most likely targets for sample return missions in the foreseeable future are small bodies (asteroids and comets), the Moon and Mars, even if one cannot exclude a return of a sample of the upper atmosphere of Venus.

References

- Albee AL, Arvidson RE, Palluconi F, and Thorpe P (2001) Overview of the Mars Global Surveyor mission. *Journal of Geophysical Research-Planets* 106: 23291–23316.
- Bate RR, Mueller DD, and White JE (1971) *Fundamentals of Astrodynamics*. New York: Dover Publications, ISBN 0-486-60061-0.
- Battin RH (1999) *An Introduction to the Mathematics and Methods of Astrodynamics, Revised edn.* Reston: AIAA, ISBN 1-56347-342-9.
- Belbruno EA and Miller J (1993) Sun-perturbed Earth-to-Moon transfers with ballistic capture. *Journal of Guidance Control and Dynamics* 16: 770–775.
- Casalino L, Colasurdo G, and Pastrone D (1998) Optimization of Delta V Earth gravity-assist trajectories. *Journal of Guidance Control and Dynamics* 21: 991–995.
- Cheng AF, Santo AG, Heeres KJ, *et al.* (1997) Near-Earth asteroid rendezvous: Mission overview. *Journal of Geophysical Research* 102: 23695–23708.
- Chesta E, Estublier D, Fallis B, *et al.* (2006) Flexible variable-specific-impulse electric propulsion systems for planetary missions. *Acta Astronautica* 59: 931–945.
- Craig S and Lyne JE (2005) Parametric study of aerocapture for missions to Venus. *Journal of Spacecraft and Rockets* 42: 1035–1038.
- D'Amario LA, Bright LE, and Wolf AA (1992) Galileo trajectory design. *Space Science Reviews* 60: 23–78.
- Donahue BB, Green SE, Coverstone VL, and Woo B (2006) Chemical and solar-electric propulsion system analyses for Mars sample return missions. *Journal of Spacecraft and Rockets* 43: 170–177.
- Edwards DL, Hubbs WS, Wertz GE, *et al.* (2004) Electron radiation effects on candidate solar sail material. *High Performance Polymers* 16: 277–288.
- Farquhar RW (1991) Halo orbits and lunar swingby missions of the 1990's. *Acta Astronautica* 24: 227–234.
- Farquhar RW (2001) The flight of ISEE-3/ICE: Origins, mission history and a legacy. *Journal of the Astronautical Sciences* 49: 23–73.
- Farquhar RW and Dunham DW (1999) The indirect launch mode: A new launch technique for interplanetary missions. *Acta Astronautica* 45: 491–497.
- Farquhar RW, Kawaguchi J, Russell C, Schwehm G, Veverka J, and Yeomans D (2002) Spacecraft exploration of asteroids: The 2001 perspective. In: Bottke W, Cellino A, Paolicchi P, and Binzel RP (eds.) *Asteroids III*, part III, pp. 367–376. Tucson: University of Arizona Press.
- French JR and Cruz MI (1980) Aerobraking and aerocapture for planetary missions. *Astronautics and Aeronautics* 18: 48–56.
- Hechler M (1997) ROSETTA mission design. *Advances in Space Research* 19: 127–136.
- Hechler M, Langevin Y, and Yanez A (2005) Mars Express orbit design evolution. *Dynamic Systems* 20: 175–188.
- Hughes GW, Macdonald M, McInnes CR, Atzei A, and Falkner P (2006) Terrestrial planets sample return mission using solar sail propulsion. *Acta Astronautica* 59: 796–806.
- Justus CG, Duvall A, Keller VW, Spilker TR, and Lockwood MK (2005) Connecting atmospheric science and atmospheric models for aerocapture at Titan and the outer planets. *Planetary and Space Science* 53: 601–605.
- Kawaguchi J (2002) A comet nucleus sample return mission using solar electric Earth gravity-assist. *Advances in Space Research* 29: 1215–1220.
- Knezevic Z and Milani A (1998) Orbit maintenance of a lunar polar orbiter. *Planetary and Space Science* 46: 1605–1611.
- Landis GA (2006) Robotic exploration of the surface and atmosphere of Venus. *Acta Astronautica* 59: 570–579.
- Langevin Y (2000) Chemical and Solar Electric Propulsion options for a cornerstone mission to Mercury. *Acta Astronautica* 47: 443–452.
- Langevin Y (2001) Solar system: Exploration. In: Murdin P (ed.) *Encyclopedia of Astronomy and Astrophysics*, pp. 2754–2762. Beograd: Institute of Physics.
- Langevin Y (2005) Design issues for space science missions. In: Martinez-Pillet V, Aparicio A, and Sanchez F (eds.) *Payload and Mission Definition in Space Sciences*, ch. 2, pp. 17–89. Cambridge: Cambridge University Press.
- Longman RW and Schneide AM (1970) Use of Jupiter moons for gravity-assist. *Journal of Spacecraft and Rockets* 7: 570–578.
- Longuski JM and Williams SN (1991) Automated design of gravity-assist trajectories to Mars and the outer planets. *Celestial Mechanics* 52: 207–220.
- Marsden JE and Ross SD (2006) New methods in celestial dynamics and mission analysis. *Bulletin of the American Mathematical Society* 43: 43–73.
- McAdams JV, Dunham DW, Farquhar RW, Taylor AH, and Williams BG (2006) Trajectory design and maneuver strategy for the Messenger mission. *Journal of Spacecraft and Rockets* 43: 1054–1064.
- McInnes CR (2003) Low cost Mercury orbiter and sample return missions using solar sail propulsion. *Aeronautical Journal* 107: 469–478.
- Peralta F and Flanagan S (1995) Cassini interplanetary trajectory design. *Control Engineering Practice* 3: 1603–1610.

- Petropoulos AE, Longuski JM, and Bonfiglio EP (2000) Trajectories to Jupiter via gravity assists from Venus, Earth and Mars. *Journal of Spacecraft and Rockets* 37: 776–783.
- Racca GD, Marini A, Stagnaro L, et al. (2002) SMART-1 mission description and development status. *Planetary and Space Science* 50: 1333–1337.
- Santo AG, Gold RE, McNutt RL, et al. (2001) The Messenger mission to Mercury: Spacecraft and mission design. *Planetary and Space Science* 49: 1481–1500.
- Scoon G, Whitcomb G, Eiden M, and Smith A (1991) Cassini-Huygens entry and descent technologies. *Space Technology* 11: 167–177.
- Shirley DL (2005) (2001) The Mariner 10 mission to Venus and Mercury. *Acta Astronautica* 53: 375–385.
- Smith JC and Bell JL (2001) Mars Odyssey aerobraking. *Journal of Spacecraft and Rockets* 42: 406–415.
- Spilker TR (2005) Significant science at Titan and Neptune from aerocapture. *Planetary and Space Science* 53: 606–616.
- Sweetser T, Peterson C, Nilsen E, and Gershman B (2003) Venus sample return missions – A range of science, a range of costs. *Acta Astronautica* 52: 165–172.
- Tsiolkovsky KE (1903) Exploration of the Universe with Reaction Machines. *The Science Review* 5.
- Tu KY, Munir MS, Mease KD, and Bayard DS (2000) Drag-based predictive tracking guidance for Mars precision landing. *Journal of Guidance Control and Dynamics* 23: 620–628.
- Vallado DA (2001) *Fundamentals of Astrodynamics and Applications*, 2nd edn. Dordrecht: Microcosm/Kluwer, ISBN 1-1881883-12-4.
- Vinh NX and Kuo ZS (1997) Improved matched asymptotic solutions for deceleration control during atmospheric entry. *Acta Astronautica* 40: 1–11.
- Willcockson WH (1999) Stardust sample return capsule design experience. *Journal of Spacecraft and Rockets* 36: 470–474.
- Winter OC (2000) The stability evolution of a family of simply periodic lunar orbits. *Planetary and Space Science* 48: 23–28.
- Wolf AA and Smith JC (1995) Design of the Cassini tour trajectory in the Saturnian system. *Control Engineering Practice* 3: 1611–1619.
- Wu SF, Costa RR, Chu QP, Mulder JA, and Ortega G (2001) Nonlinear dynamic modeling and simulation of an atmospheric re-entry spacecraft. *Aerospace Science and Technology* 5: 365–381.
- Yarnoz D, Jehn R, and Croon M (2006) Interplanetary navigation along the low-thrust trajectory of BepiColombo. *Acta Astronautica* 59: 284–293.
- Yen CS (1989) Ballistic Mercury orbiter mission via Venus gravity-assist. *Journal of the Astronautical Sciences* 37: 417–432.

10.18 Instrumentation for Planetary Exploration Missions

P. Falkner, A. Peacock, and R. Schulz, European Space Agency, Noordwijk, The Netherlands

© 2007 Elsevier B.V. All rights reserved.

10.18.1	Introduction	597
10.18.2	Building on Past Missions	598
10.18.3	Exploration Strategies and Associated Techniques	599
10.18.3.1	Exploration Strategies for Near-Future Perspectives	599
10.18.3.1.1	Exploration of the Jovian system	599
10.18.3.1.2	Preparation of landing on terrestrial planets	600
10.18.3.1.3	Sample and return from a low-gravity environment	601
10.18.3.2	Longer-Term Goals for Exploration	601
10.18.4	Instruments	602
10.18.4.1	Measurement Principles	602
10.18.4.1.1	Direct detection	602
10.18.4.1.2	Indirect detection	602
10.18.4.1.3	Remote sensing	602
10.18.4.1.4	<i>In situ</i> sensing	603
10.18.4.1.5	Passive detection	603
10.18.4.1.6	Active detection	603
10.18.4.2	Imager	604
10.18.4.2.1	Wide field imager	604
10.18.4.2.2	High-resolution camera	605
10.18.4.2.3	Stereo imager	605
10.18.4.2.4	Scanning platform	605
10.18.4.2.5	Panoramic camera	605
10.18.4.2.6	Microscopic Imager	606
10.18.4.3	Spectrometer	606
10.18.4.3.1	Visual and Infrared Mapping Spectrometer	607
10.18.4.3.2	UV spectrometer	607
10.18.4.4	Radiometer and Polarimeter	608
10.18.4.4.1	Thermal IR radiometers	608
10.18.4.5	Nephelometer	608
10.18.4.6	X-Ray Spectrometer	609
10.18.4.6.1	X-ray fluorescence	609
10.18.4.6.2	Particle-induced X-ray emission	609
10.18.4.6.3	Alpha particle X-ray spectrometer	609
10.18.4.7	Gamma-Ray Spectrometer	610
10.18.4.8	Neutron Spectrometer	610
10.18.4.9	Laser Altimeter	611
10.18.4.10	Radar	612
10.18.4.10.1	Radar altimeter	612
10.18.4.10.2	Ground penetrating radar	612
10.18.4.10.3	Synthetic aperture radar	612
10.18.4.11	Radio Science	614
10.18.4.11.1	Celestial mechanics and the gravitational field	614
10.18.4.11.2	Radio propagation at occultation	614
10.18.4.12	Microwave Sounder	615
10.18.4.13	Particle Detector	615

10.18.4.13.1	Heavy ion counter	615
10.18.4.13.2	Energetic particle analyzer	615
10.18.4.13.3	Neutral particle analyzer	615
10.18.4.14	Mass Spectrometer	616
10.18.4.14.1	Magnetic focusing analyzer	616
10.18.4.14.2	Quadrupole mass analyzer	616
10.18.4.14.3	Ion trap	617
10.18.4.14.4	Wien filter	617
10.18.4.14.5	Gas chromatograph mass spectrometer	618
10.18.4.14.6	TOF mass analyzer	618
10.18.4.14.7	Laser ablation mass spectrometer	618
10.18.4.14.8	Laser-induced breakdown spectroscopy	618
10.18.4.14.9	Raman spectrometer	619
10.18.4.15	Mössbauer Spectrometer	619
10.18.4.16	Dust Analyzer	620
10.18.4.17	Radiation Environment Monitor (REM)	620
10.18.4.18	Scanning Tunneling and Atomic-Force Microscope	620
10.18.4.19	Scanning Electron Microscope	621
10.18.4.20	Plasma and Wave Instruments	621
10.18.4.20.1	Plasma wave analyzer	621
10.18.4.20.2	Plasma analyzer	622
10.18.4.20.3	Langmuir probe	622
10.18.4.20.4	Electron drift instrument	622
10.18.4.20.5	Active spacecraft potential control	623
10.18.4.21	Magnetometer	623
10.18.4.21.1	Fluxgate magnetometer	623
10.18.4.21.2	Vector helium magnetometer	623
10.18.4.21.3	Search coil	624
10.18.4.22	Accelerometer and Seismometer	624
10.18.4.22.1	Seismometer	625
10.18.4.23	Thermometer	625
10.18.4.24	Pressure	625
10.18.4.25	Micro-Gas Sensors	626
10.18.4.26	Surface Dating	626
10.18.5	Instrument Suites for Exploration	627
10.18.5.1	Planetary Size, Mass, Shape, and Interior	628
10.18.5.2	Surface and Subsurface	628
10.18.5.2.1	Surface station	629
10.18.5.2.2	Rover	629
10.18.5.2.3	Subsurface access	629
10.18.5.2.4	Mechanical tools	630
10.18.5.3	Atmospheres, Exospheres, and Ionospheres	630
10.18.5.3.1	Entry and descent probes	631
10.18.5.3.2	Aerobots	631
10.18.5.4	Magnetosphere	632
10.18.5.5	Moons	633
10.18.5.6	Rings	633
10.18.5.7	Asteroids and Comets	634
10.18.5.8	Exobiology	636
10.18.5.9	Sample Return	636

10.18.5.10	Engineering Aspects and Operations	637
10.18.5.10.1	Engineering constraints	637
10.18.5.10.2	Operational aspects	637
10.18.5.10.3	Verification and testing	638
10.18.6	Outlook and Timeline	638
References		638

10.18.1 Introduction

The bodies making up our solar system are usually divided into different groups (*see* Chapter 10.01). Often we simply distinguish between inner and outer planets, moons, and small bodies, namely comets and asteroids. However, if we wish to find ways to investigate these bodies through planetary exploration missions, we need to first identify those of their basic properties that are relevant for the definition of the mission. Different solar system objects need different study strategies and therefore different types of instrumentation. Specifically, we have to distinguish between the measurements needed to understand:

- terrestrial planets and large moons with atmosphere;
- terrestrial planets and large moons without atmosphere;
- giant planets and their environment, for example, Jovian magnetosphere, Io torus, Saturn rings;
- dwarf planets and trans-Neptunian objects (TNOs);
- small bodies without atmosphere, for example, asteroids, small moons;
- comets; and
- interstellar and interplanetary dust particles.

Apart from the obvious need to design an exploration mission to any of these objects in accordance with the conditions expected at the object's location (e.g., thermal and radiation environment), the mission scenario and scientific instrumentation need to be specifically tailored to the properties of the object itself. For instance, the requirements are very different for landing on a terrestrial planet or a small body with low gravity. Similarly, the requirements on exploring an airless body are very different from those required for a body with a thick atmosphere. In addition, the selection of the mission scenario depends decisively on what we already know of the object(s) from other investigations and past missions. For example, if a certain object has been visited and studied by remote-sensing techniques from orbit to

an extend that no epoch-making results can be expected from another remote-sensing mission, the logical next step should be to obtain *in situ* measurements. Depending on the type of body, the *in situ* measurements would involve the study of the atmosphere, surface/subsurface, or both.

Generally, one can distinguish between three broad classes of measurements, namely remote, environmental, and *in situ* sensing. For missions to solar system bodies (planets, moons, asteroids, comets), the remote sensing would usually be the first step of any space investigation. Remote sensing from a spacecraft largely benefits from the simple fact that the measurements can be obtained very close to the object to be investigated, rather than from observatories on Earth or in Earth orbit. Typical remote-sensing instruments would cover the classical measurement triangle: spatial (imager)/spectral (spectrometer)/intensity (radiometer). In addition, any mission focusing on remote sensing of a solar system body from orbit or during a fly-by would also include a number of instruments for environmental sensing of any existing plasma or magnetosphere around the object. For the next step, *in situ* measurements, the choice of mission scenario and payload depends critically on the object to be investigated, as well as on which aspect of this object shall be studied. For example, if we wish to perform *in situ* measurements of the atmosphere of a body, the requirements for the mission will be very different from those for *in situ* investigations of the (sub-) surface composition. For the latter, it is mandatory to actually land on the object under investigation, while the former might for some objects (comets, asteroids, planets/moons with exosphere) be accomplished by simply adding to the payload instruments with *in situ* measurement capabilities, such as for instance mass spectrometers. In other cases, particularly for bodies with dense atmospheres, such as Venus, Titan, and the gas giants, an atmospheric entry probe is required.

An overview is provided on possible instrumentation for various types of planetary exploration missions adjusted to the strategies required for the respective investigations.

10.18.2 Building on Past Missions

Spacecraft have visited each of the eight planets of our solar system, as well as a number of moons, comets, and asteroids. The levels at which certain objects have been investigated go hand in hand with the types of missions that have been conducted so far. At present, samples from the Moon and cometary dust have been returned to Earth. The Stardust Mission has collected dust particles from the coma of comet 81P/Wild 2. The sample return capsule with the dust particles stowed into aerogel was safely returned to the Earth on 15 January 2006 (Brownlee *et al.*, 2006).

Few bodies have been investigated *in situ*. The thick atmospheres of Venus, Jupiter, and Titan were studied through atmospheric probes, whereas the thin but spatially hugely extended atmospheres of comets were investigated *in situ* by selected instruments capable of performing *in situ* measurements on fly-by missions. For Venus and Titan also the surface could be reached. The first landing on Venus took place in 1975 (Venera 9 and 10). Almost 30 years later the Huygens Probe successfully reached the surface of Titan to return the first images of this unusual moon. After the first landing on Mars in 1976 (Viking 1 and 2) the up-to-then tremendous interest in this planet (21 missions in 15 years) stagnated for about 10 years. Subsequently, Mars investigations suffered severely from a number of mission failures (six of nine missions failed between 1988 and 1999). However, Mars has become a prime target for comparative planetology and the search for the roots of life, and Mars exploration is nowadays regarded as one of the key elements in solar system studies (e.g., Hauber and Neukum, 2006).

The exploration of any solar system body usually includes investigations with global coverage, as well as dedicated *in situ* analysis and sample return from specific regions of particular importance. This general scheme is normally being followed for the first ‘round’ of space missions to a solar system body, although owing to technological constraints their exploration has practically always started with fly-by missions before global characterization from an orbiting spacecraft could be achieved. At present, the only object for which all the consecutive steps mentioned above have actually been performed is the Moon. However, even though several hundred kilograms of Moon matter have been returned to Earth for in-depth analysis, we recently have begun a return to the Moon for global mapping, even without

landing, some 35 years after the lunar sample returns. Why? Actually we are going back for various reasons. First, it has to be verified whether the results of the detailed physical and chemical analysis of the lunar samples in the laboratory indeed apply globally to the Moon and do not just reflect some odd behavior of the particular regions, for which samples were returned. Second, a lot of progress was made over the past 35 years in modeling the formation and evolution of solar system bodies in terms of their accretion and subsequent alteration through endogenic and exogenic processes. These models have not yet been verified through detailed measurements, largely because appropriate instruments were not available in the 1970s. Finally, a lot of new knowledge has been gathered in recent years concerning for instance the presence of volatiles in polar regions. All these issues can only be addressed by returning to the Moon with innovative instrumentation. To fully understand how the existence of the Moon constrains our models for the formation and evolution of the terrestrial planets, we need knowledge on its global composition, interior structure, and thermal properties, such as for instance the surface heat flow. Last, but not the least, owing to its small distance to Earth, the Moon is ideal to be used as test bed for exploration of other planets (e.g., Mars).

By now, Venus, Mars, Jupiter, and Saturn have been explored from orbiting satellites; Uranus, Neptune, and Mercury were investigated in part through fly-by missions, while Pluto is still awaiting its first encounter. Missions are on their way to investigate Mercury from orbit and to orbit and land on a comet nucleus. While we have had a number of asteroid and comet fly-bys, so far a close study from orbit has only been taken on Eros and on Itokawa by the Near Earth Asteroid Rendezvous (NEAR) and Hayabusa Missions, respectively.

Of course, we all understand that due to financial constraints only a very limited number of space missions can be conducted at all in the next few decades. It is therefore of utmost importance to plan these mission well ahead in terms of which targets need to be visited, in which time frame, and what measurements have to be obtained of any specific target in the near and long-term future. To be able to take these important decisions one must review the results of past missions, with a view to establishing which questions have been answered and which have remained unanswered or have not yet been addressed and why. We then have to find the appropriate strategy with a view to covering as many of the important open

issues as possible within the financial and technological constraints. This approach highlights in many cases that the scientific return from past missions simply reflects the technology limitations of their time. This is particularly valid with respect to sensitivity, as well as spatial and spectral coverage of remote-sensing instrument and the availability of *in situ* measurements, suggesting that progress in scientific research goes hand in hand with the progress made in the technology of scientific instrumentation, at least for space missions.

10.18.3 Exploration Strategies and Associated Techniques

As already mentioned, the exploration of the different types of solar system objects requires different mission strategies, as well as different instrumentations. This however does not mean that a dedicated mission is required every time to one target only. Missions have been designed to combine whenever possible investigations of several objects and to cover various aspects of the chosen targets. The Galileo Mission flew past two asteroids before it had its encounter with Jupiter, and Rosetta will do the same before encountering its target comet. The most famous example of taking advantage of rare, but extremely fortunate geometrical constellations in the solar system is Voyager 2 using the once-every-189-years alignment of the outer planets, which allowed this spacecraft to visit all outer giant planets. However, although science can greatly benefit from spacecraft visiting various targets by the side, the mission and payload was always designed for the investigation of the main target, and in most cases the additional scientific return was mainly limited to images of any additionally visited body.

Future exploration strategies should therefore take advantage not only of fortunate geometrical circumstances, but also of advances in technology of instrumentation, which includes sensor development, as well as low-resource approaches for support electronics. In the following sections some considerations are presented for shorter-term and longer-term strategies for the exploration of the inner and outer solar system.

10.18.3.1 Exploration Strategies for Near-Future Perspectives

At this point in time, planetary space exploration includes missions to Saturn and Titan

(Cassini/Huygens), Mars (Mars Express, Mars Reconnaissance Orbiter), Venus (Venus Express), Mercury (Messenger and BepiColombo), the Moon (Smart-1, Chandrayan, Chang'E), as well as various comets (Rosetta, Stardust) and asteroids (Hayabusa, Rosetta). Some of these missions actually represent the realization of the ‘next step’ in a strategic exploration approach as defined above. For instance, Rosetta represents the ‘next step’ after cometary fly-by missions; Cassini represents the return to the Saturn system after Voyager, with dedicated payload and with the Huygens Probe for detailed *in situ* investigations of the Titan atmosphere. When reviewing the current and past missions in relation to the status of a repeated four-step exploration approach as outlined in Section 10.18.2 (fly-by, orbiter, lander, sample return), it becomes clear that for most targets (Mars, Venus, Mercury, asteroids, and comets) in the near future, missions will need to concentrate on further *in situ* investigations from landers (atmospheric entry probes in case of Venus) and the preparation of sample return missions. After Voyager and Galileo, it is obvious that we need to return and explore the Jovian system to obtain a clear understanding of Jupiter and its magnetosphere, as well as the Galilean moons. At some point a decision is needed on which targets to concentrate on and what type of missions to design. Three scenarios are in principle competing candidates for near-term future missions:

- exploration of the Jovian system,
- landing on terrestrial planets, and
- sample and return from low-gravity environment.

We will briefly introduce possible strategies for these three types of planetary missions.

10.18.3.1.1 Exploration of the Jovian system

A staggered approach is envisaged with the final goal of having explored the major aspects of the Jovian system in a detail that allows us to understand the formation and evolution of the system as a whole, as well as of its individual components. Owing to the hostile environment, particularly with regard to radiation, a dual-spacecraft approach is proposed for each individual step of exploration. One spacecraft would be used for the dedicated study of the prime target, which for the first step could be Jupiter or a Galilean moon, while the second spacecraft would function as the relay satellite, performing command and control functions with some additional scientific capabilities. The Jovian Relay Satellite

could be placed in a highly elliptical orbit around Jupiter, outside its main radiation belts, and would therefore be able to relay science data to Earth for approximately 2 years. To reach the right ratio between scientific merit and cost of the mission, both spacecraft have to be small enough to be launched economically and at the same time should carry a sophisticated, highly miniaturized and integrated payload, meeting the science requirements of the mission. Here as an example we focus on an orbiter mission to a Galilean moon, for which we take Europa as an example, as well as the study of the Jovian magnetosphere and *in situ* studies of Jupiter's atmosphere.

The lifetime of a low-resource Europa Orbiter is limited to approximately 60 days once in Europa orbit owing to the extremely harsh radiation environment. At the same time the data rate over 5.2 AU only allow for the direct transmission of a limited amount of data. Therefore, this spacecraft will not nearly be able to uplink all the scientific data it will accumulate to Earth directly within its mission. The use of the relay spacecraft at a relatively close distance will ensure that the data can be relayed within the specified lifetime of 60 days. By placing the relay satellite outside the main radiation belts, its lifetime will be more than sufficient to allow the download of all received data to Earth. Furthermore, the relay satellite would be able to also carry a certain amount of scientific payload, to perform for instance measurements of the Jovian magnetosphere, for which it could also go into a more suitable orbit after the Europa Mission is terminated. The relay satellite could also be used to perform remote-sensing studies of Jupiter's atmosphere. The remote-sensing instruments will however require an orbit that is close enough to obtain sufficiently high resolution. If the orbit is however lowered to inside the main radiation belts, the mission lifetime will be severely reduced and the instrumentation performance could be significantly degraded.

A very interesting scientific contribution would be provided by an entry probe, capable of penetrating the denser layers of the atmosphere, if possible up to a pressure of 100 bar. The Galileo probe entered a dry region of the Jovian atmosphere, preventing the assessment of critical key measurements such as representative O/H ratios. To avoid duplication, a different entry strategy is required, such as, for instance, entering at higher latitude, and preferably with multiple probes. The orbits required for a high-latitude entry can however result in higher approach

velocities, hence increase the entry velocity of the probe. This will increase the heat load on the thermal protection structure, which is already stretched to the limit of current capabilities for equatorial entry of the Jovian atmosphere. Different deployment strategies have to be assessed: from a magnetospheric spacecraft the probe would be released either from hyperbolic approach or from Jovian orbit. Should a strong case be made for multiple probes, it becomes highly unlikely that a relatively small magnetospheric spacecraft can carry the probes, and in that case a dedicated system will have to be assessed.

10.18.3.1.2 Preparation of landing on terrestrial planets

Depending on the presence and the thickness of their atmospheres the landing on the terrestrial planets requires different strategies. For atmosphereless bodies such as Mercury and the Moon the decent and soft landing must rely solely on propulsion systems and airbags. For landing on Mars and Venus, on the other hand, the descent goes through the atmosphere, which requires an atmospheric entry shield. Closer to the surface the deceleration process can be supported by parachutes and the touchdown by airbag systems.

Mars has become the target of prime interest for landing and *in situ* exploration. In preparation for a landing mission to Mars, two aspects need to be considered in parallel. (1) A number of improvements of the existing entry decent and landing systems will secure the safe landing on Mars, including autonomous hazard avoidance and pinpoint landing. Improved accuracies of the eject mechanism and optimization of the parachutes and airbags will further improve the safety of the landing. (2) As the landing on Mars shall be the first step toward a sample return mission, it is of utmost importance that a large area around the landing site is studied. A rover equipped with a deep drill and with science packages that include biochemical and geophysical instrumentation is required.

The Moon would be an excellent test bed for the velocity-controlled pinpoint landing system and also for the performance verification of the mobile elements and science packages. Operational sequences could be tested and improved. It has the major advantage of being very close to Earth, which reduces the traveltimes (and therefore cost) and simplifies communication and commanding. A landing system leading to a safe pinpoint landing on the Moon can

provide a solid heritage for pinpoint landing on Mars, of course apart from the additional entry system and the parachutes, which decrease the required thrust. The mobile surface elements and science packages can be fully tested on the Moon before being sent to Mars.

10.18.3.1.3 Sample and return from a low-gravity environment

Small solar system bodies, such as comets, asteroids, and TNOs, are believed to be relics of the formation of our planetary system. The near-Earth objects (NEOs) are very attractive targets for sample return missions, because the technological challenges are less demanding than for the high-gravity environment of a planet or the volatile composition of a comet. NEOs can be reached on much shorter time-scales than Trojan asteroids, Centaurs, or TNOs. Dynamical calculations have shown that their orbital lifetime is small on cosmogonical timescales and that they are continuously being replenished from other sources, either from the main asteroids belt or from comets. A sample return mission needs to return a scientifically significant sample of the surface material of one or multiple NEOs. The amount of material will influence the research that can be performed. It would be advantageous to have enough material to apply all desired measurement techniques and tests. Most laboratory instruments require a very small amount of material ($<<1\text{ g}$) and only some investigations need a few grams. The sampling size required for one set of measurements is therefore of the order of several tens of grams. However, some redundancy is required and some additional material should be kept for future analysis. Furthermore, a larger amount would provide a better representation of the sampled area.

Retrieving a sample from a low-gravity body is significantly different from a sample return from a planet. The latter requires a lander, an entry and/or descent system, and a launcher to return to orbit. For a low-gravity environment, no dedicated launch vehicle is required after sample collection, and also the descent requirements are very different. A sample return from a NEO, a main belt asteroid or one of the Martian moons, can be achieved with a small spacecraft that can be launched on a ‘low cost’ launcher, if the relative velocity between the spacecraft and the target body is low enough to permit a rendezvous mission. Upon arrival the spacecraft will first go in orbit to perform a number of remote-sensing

measurements of the surface and subsurface, as well as an accurate mapping of its gravitational properties, which will be a key prerequisite in the detailed determination of the navigation sequence for sampling maneuvers. A large number of potential sampling sites on the surface need to be examined in order to select the optimal locations for sampling maneuvers. For this high-resolution imaging in the visual and near infrared (IR) is the minimum requirement, however, depending on the available mass other instruments should be added to determine the global elemental and mineralogical surface composition of the target, as well as some information on the subsurface.

There are two main options to retrieve a sample from a low-gravity solar system body: (1) collecting the sample directly from the surface or (2) creating a debris cloud of asteroid material and collecting a sample from that cloud. The Hayabusa Mission to asteroid Itokawa has realized the first option whereby a touch-and-go sampling approach was used. Upon a brief contact with the surface, a projectile is fired and the debris is then funneled up a cone as the spacecraft retreats from the surface. The second option has not yet been space proven; however, the results of the Deep Impact Mission to comet Temple 1 can be used to roughly estimate the constraints for collecting a sample of some hundred grams with this method from a nonvolatile asteroid.

The samples have to be stored and sealed into a sample canister inside an Earth-entry vehicle. To reduce propellant requirements for the transfer back to Earth, the sampling mechanism and empty tanks can then be jettisoned. The spacecraft will return to Earth, where the Earth-entry capsule will perform a direct hyperbolic Earth entry.

10.18.3.2 Longer-Term Goals for Exploration

The high-priority goal is of course a sample and return from a specific set of locations which have been previously studied by *in situ* landers or their associated rovers. While there are many differences between a sample and return mission to Mars and that to a NEO, some technologies such as the sample encapsulation and Earth-entry vehicle may be similar. Certainly, a mission to a NEO may be a natural test-bed precursor to this important further step on Mars.

10.18.4 Instruments

The challenge of any scientific mission is to define the scientific objectives (what is of highest priority to be observed), refine them in detailed scientific requirements, and select an instrument or instrument combination to satisfy these requirements. In most of the cases the instrument accommodation drives the spacecraft design and hence strongly influences the overall cost of the mission. Several iterations are usually required to fit the actual selection and implementation details within the available resources. Tradeoffs on the type of instruments, the instrument measurement principle, and combination of instruments and level of new technology are needed. Usually, many compromises have to be made in the instrument selection and accommodation, together with the verification that the scientific objectives of the mission are still fulfilled. Several textbooks are dealing with this space-mission design approach (e.g., Wertz and Larson, 1999) (Figure 1).

To satisfy the objectives of the mission a set of physical qualities has to be measured and observed. The underlying physical principle allows for various observation methods. A classification of instruments according to these measurement principles should assist the reader in obtaining a structured overview of the types of instrumentation required for planetary space missions.

10.18.4.1 Measurement Principles

Planetary exploration covers a wide area from astronomy, geology, morphology, geophysics over magnetospheric and plasma science, to exobiology, just to name, but a few of the involved disciplines. A wide and ever-growing set of instruments are employed driven by new measurement requirements and new scientific disciplines on the one side and emerging new technology on the other side, enabling enhanced and new instruments. A classification of detection principles should assist in categorizing the existing range of instruments based on their measurement approach. We distinguish between direct and

indirect detection, remote and *in situ* sensing, passive and active sensing principles.

10.18.4.1.1 Direct detection

The detection principle is based on a direct measurement of the interesting physical phenomena by observation of physical quantity such as temperature (thermal sensor), pressure (pressure sensor), and particle flux by means of either remote sensing or *in situ* observation.

10.18.4.1.2 Indirect detection

Indirect detection is performed when a direct detection is not possible or too demanding. The observation of the interesting phenomena is done by interpretation of a single or combination of several physical byproducts which are easier to access for observation than a direct measurement. Indirect detection often relies on a combination of measurements performed with a set of instruments and very often involves modeling and simulation to derive the physical result of interest. Many processes and physical phenomena have to be measured indirectly because direct access is not possible due to demanding or inadequate space-qualified sensor systems.

Tracing of Radon as product of radioactive decay of uranium-238 to conclude on volcanic activity on planetary surfaces is an example of an indirect observation of geophysical phenomena (although in many cases volcanic activity could be detected directly by imaging, e.g., provided that visibility is given and no opaque atmosphere like on Venus or Titan blocks the direct observation).

10.18.4.1.3 Remote sensing

Any observation without directly contacting the object in question is considered remote sensing. Mainly electromagnetic waves allow for remote detection of physical qualities either directly or indirectly. The detection of electromagnetic waves covers a wide wave-band range (submillimeters (terahertz), thermal IR, far and near IR, IR, visible, ultraviolet (UV), extended UV, to X-ray and gamma rays). The emission of photons from a planetary surface allows for the characterization, definition of chemical composition, morphology, mineralogy, etc. Due to the propagation of these photons it is possible to detect them remotely without direct interaction of an instrument at the origin or location of the phenomena. This is exploited with either remote observation from Earth (as far as the atmospheric attenuation allows) or outside the Earth atmosphere

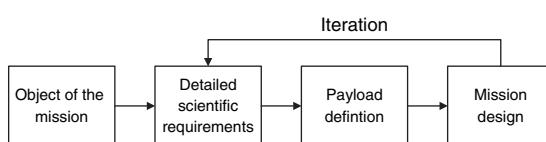


Figure 1 Definition process of a space mission.

with space telescopes (Hubble, JWST) or in closer vicinity with remote-sensing spacecrafts (fly-bys or in orbit). A major limiting factor for remote sensing is the distance from the observed phenomena due to the $1/r^2$ dependence from distance resulting in a limited detectable signal strength and signal-over-noise ratio (SNR). Spatial and spectral resolution is restricted both by the photon wavelength, as well as the limited instrument apertures (size, mass, and cost of an instrument), and available signal strength. Of course, the benefit of remote sensing is that in many cases a wide coverage on an object can be achieved. Examples of remote-sensing instruments are (telescopic) imagers, laser altimeters, or radar.

10.18.4.1.4 In situ sensing

Being at the place of a physical phenomena (e.g., flying or floating through a planetary atmosphere with an aerobot or being on a planetary surface with lander and rovers) allows measurement of the physical qualities directly at the location. *In situ* sensing is the only possible means to investigate phenomena which have no emission of or interaction with electromagnetic waves (or where the propagation of waves is suppressed or strongly attenuated). Due to the vicinity very often much-lower intensity or less-disturbed signals can be measured. In most cases *in situ* observations are of local nature (unless a high level of mobility is provided by long-distance rovers, aerobots, or orbiting spacecraft), leading to a restricted coverage. Nevertheless, it is important to provide ground truth for remote-sensing measurements, which rely often on modeling and simulations (required due to indirect measurement principle) to derive the desired information. A further benefit of *in situ* instrumentation is the possible physical interaction with the object, for example, by use of mechanical means such as grinders or drill, allowing for access of pristine and ‘unweathered’ material below the surface.

A major drawback for most *in situ* instrumentation is the increased complexity of the required equipment, as the instruments have to be brought to the targeted location under specific, often very challenging environmental conditions (e.g., radiation at Jupiter or extreme temperatures at Venus and Mercury). In many cases, *in situ* observations drive the mission design and increase the mission complexity, as the instruments cannot remain at a remote location but have to be brought by landers, aerobots, impactors, etc., to the location of interest.

Some examples of *in situ* instruments are dust mass spectrometers, plasma analyzers, laser mass spectrometers, or magnetometers. Note the return of a sample from a planetary surface (such as during the lunar Apollo program, Genesis, or as foreseen for Hayabusa) also belongs in principle to the *in situ* observation category.

10.18.4.1.5 Passive detection

Passive instruments do not provide their own illumination or stimulation, but they relay on either illumination from natural sources (e.g., by the Sun) or luminescence, fluorescence, or emission of radiation. The instrument does not need illumination capabilities, but is directly dependent on the illumination or emission source.

Major parts of the solar emission spectra are reflected from planetary surfaces and altered by the surface interaction. A relative comparison of the reflected spectra with the original impinging solar spectrum reveals many surface properties. In the case of variable illumination (as in the case of solar illumination, but dependent on the wave band), a reference measurement is required to derive the relative ratio caused by the surface interaction, whenever absolute information is required. Some examples of passive instruments are microwave sounders, radiometers, X-ray spectrometers (could be also active), or magnetometers.

10.18.4.1.6 Active detection

Here the instrument provides its own illumination using an emitter. A signal source is used to illuminate the region of interest (footprint). The sensing part is co-aligned with the illuminating beam and detects the reflected signal. The benefit is of course the independence from other sources of illumination (e.g., use on the night side of a planet) due to the well-known emitted signal characteristics (time of emission, spectral features, power, etc.). The modulation of the emitted signal (e.g., with pseudo-codes) can dramatically increase the SNR and assist in the detection of the emitted signal even under noisy conditions. No direct dependence on, for example, solar aspect angle (illumination aspect) or solar variability is required with this technique. The drawback is the increased complexity and resource requirements for active illumination (mainly in mass, volume, and power). It is not necessarily the case that the emitter must be in close vicinity of the receiver, but can be on another spacecraft, lander,

or even located on Earth. Some examples of active instruments are radars, laser altimeters, permittivity probes, or laser mass spectrometers.

10.18.4.2 Imager

Imaging is used to provide spatial information about the distribution of any interesting physical quality. Imaging can be achieved either by repeated measurements at different targets (e.g., with a scanning principle like push broom and whisk broom) using a single detector or by simultaneous parallel measurements by means of a detector array employing multiple sensors synchronously. A scanning principle requires a mechanical moving platform or relies on the motion of the instrument (e.g., mounted on an orbiting spacecraft). Scanning is used in cases where the detection principle is complicated and a complex and demanding instrument setup is required and also in cases where the required pixel number is exceeding the available detector array sizes, or high-resolution spectral information is required. The imaging quality is subject to the pointing accuracy and stability of the mechanical scanning platform or the orbiting spacecraft.

In many cases multiple detectors can be used to perform simultaneous measurements, simplifying or even avoiding the scanning platform. An optic is required to produce an image on the detector array. Very often individual images have to be put together to provide a global image of a large spatial area. This can become rather complicated if observation distances, aspect angels, and illumination conditions are varying between individual images.

Imaging is very important to recognize features, structures, surface morphology, and context for other instruments or simply for orientation and navigation purposes. An imaging instrument uses optics such as lenses or mirrors to project a focused image onto a detector. The detector converts the image into electrical signals, which are digitized before they are further processed with a digital data processing unit. Color imaging requires multiple exposures of the same target through different color filters. The combination of three color-filtered images in complementary colors such as red, yellow, and blue provide natural colored images. In several applications filter wheels have been used to place the required filter in the optical path. More recent developments however make use of interference filters applied directly on the detector array, which does, in principle, save mass. False color images are used to

highlight mineralogical composition or other specific features within the image.

Imagers consist typically of a solid-state array based on either charge-coupled device (CCD) or an active pixel sensor (APS). In earlier times also direct recording on film material was used.

Light falling on a detector pixel is absorbed by a photoconductive substrate such as silicon and releases a quantity of electrons proportional to the intensity of light. The stored accumulated electrical charge in a pixel represents the light level integrated over the exposure time. These charges are read out for conversion into digital data. Typical CCD and APS are sensitive over a wide wave band (including UV and IR). Their sensitivity can also be increased to enable the detection of single photons at specific wavelengths.

Not all imagers require two-dimensional (2-D) detector arrays. Imaging can be done also with a single-element whisk-broom sensor, multi-element whisk-broom sensor, sensor line in a push-broom arrangement, or in matrix configuration. A single line detector arranged perpendicular to the spacecraft velocity vector can be used in a scanning mode (push-broom mode). A 2-D image is built up as the spacecraft moves around its orbit, requiring appropriate spacecraft attitude stability (to avoid image distortion during the scan) and precise pointing knowledge. An optical correlator system can significantly reduce stability requirements for the satellite attitude system.

The simplest approach for image sensing is through a detector array (typically 1024 times 1024 or larger), where a complete image is obtained in a single 'shot'. The stability requirement is now driven by the exposure time, as the platform has to remain stable during the exposure time within one pixel. The imager optics is either of a reflecting (e.g., Schmidt, Cassegrain, or Triple Mirror Anastigmat (TMA)) or refractive type (teleoptic, lens triplets).

10.18.4.2.1 Wide field imager

A wide field imager (WFI) has a limited ground pixel resolution, but gives a good overview over a larger terrain and hence is used to provide context information for other instruments such as high-resolution cameras or spectrometers. WFIs can be rather small as the larger ground pixel size allows for a reduced aperture and reduced focal length. WFIs are usually also used in connection with stereoscopic imaging. One of the best-known WFIs is the Wide Field

Planetary Camera 2 (WFPC2), which was installed at the Hubble Space Telescope in 1993 (Peg, 1993).

10.18.4.2.2 High-resolution camera

High-resolution imagers are used on remote-sensing spacecraft to resolve small details (typically in the meter range or even better) on the observed body surface. In combination with filters some high-resolution compositional information can be provided as the typically employed detector arrays (CCD or APS) are sensitive also in the IR and UV region, provided the natural illumination of the target provides enough photons per pixel in a small waveband, since the ground speed of the spacecraft limits the exposure time.

CCDs provide an advantage in the case of low photon counts as they have higher quantum efficiency than APS, although more recent developments in this area indicate that APS sensors can also achieve high efficiency. One significant benefit of CCDs is that a time delay integration (TDI) mode can be easily implemented due to their construction principle. The exposure time can be increased by shifting the pixel information with constant speed from one pixel to the adjacent pixel by appropriate clocking of the charge transport mechanism of the CCD against the projected spacecraft velocity and allow so for accumulation of photoelectric charges. If the shift is chosen in the right speed (determined by the ground track velocity of the spacecraft and the focal length of the optics), the spacecraft motion can be fully compensated and hence the exposure time increased until either saturation is achieved or the edge of the CCD is reached. Every individual ground pixel is shifted through a defined number of CCD cells, before readout.

To overcome diffraction limits the aperture of the optics has to be adapted to the required spatial resolution. The overall mass of the instrument optics unfortunately scales with the required aperture and focal length and is hence more demanding than the WFI. The spacecraft attitude control subsystem has to provide appropriate pointing stability and absolute pointing capabilities to avoid blurred and smeared images and finally a huge amount of data has to be handled by the spacecraft data handling and the data telemetry subsystems. Usually the high-resolution channel is combined with a WFI, which provides context information. Due to the amount of provided information high-resolution imaging requires a high data rate, which is usually limited. Image compression has to be used carefully to avoid artifacts in the

final data product. The High Resolution Stereo Camera (HRSC) on Mars Express (Neukum *et al.*, 2004) has a ground resolution of up to 2 m pixel⁻¹.

10.18.4.2.3 Stereo imager

Stereo imaging is used to gain topographical information and to produce digital elevation models of planetary surfaces. This can be done with either two camera heads with a spatial separation but pointing at the same target, or with a single camera head, which images a specific area and after a revisit (spacecraft flies again over the same local area) images the same area under a different aspect angle (achieved by side viewing of the spacecraft). The first case provides stereo information even with single site visits and without spacecraft off-nadir pointing, but the achievable height resolution is limited because usually the spatial separation of the camera heads is limited. The second case provides almost any required separation, but needs multiple revisits for stereo imaging and phases of spacecraft off-nadir pointing. A good compromise between both cases is the use of a second camera which is mounted off-nadir pointing.

In all cases complex image processing is required to retrieve the elevation of the terrain and to create 3-D images. Cameras with stereo capabilities are on board the orbiters of most planetary missions with the currently most impressive one being the HRSC on Mars Express.

10.18.4.2.4 Scanning platform

Optical instruments are sometimes installed on a mechanical scan platform, which points in commanded directions, allowing optical observations to be taken independently of the spacecraft's attitude. However, most modern spacecraft designs do not include such scan platforms in order to avoid moving mechanisms which are potential for failure. The alternative is to rotate the entire spacecraft so as to point the optical instrument in the required direction. This is particularly easily done with three-axis stabilized spacecraft which use reaction wheels for their attitude control, as the re-pointing can be performed without propellant use (neglecting the wheel offloading, which is required from time to time to keep the rotational rate of the reaction wheels within a specified range).

10.18.4.2.5 Panoramic camera

A panoramic camera is a high-resolution stereo imager used to image the surface and surroundings from a lander or rover. Two camera heads are spatially

separated and typically mounted on top of a mast, allowing for an elevated view. A mechanism with two degrees of freedom allows the cameras to rotate by 360° azimuth to obtain a panoramic view and provides pointing at various elevations.

Due to the spatial separation combinations of the images can be used to create 3-D maps of the area where, for example, the rover has to navigate. Mechanical filter wheels in the optical path, fixed linear or interference filters directly mounted on the CCD or APS gives multispectral imaging capabilities and offers color imaging. Images taken at various wavelengths are used to study mineralogical composition of interesting rocks and soils. Examples for such systems are on the Mars rovers Pathfinder, Spirit, and Opportunity (e.g., Bell *et al.*, 2006), as well as on the Rosetta Lander Philae (e.g. The Philae Team *et al.*, 2006).

A difficulty however arises from the large volume of data and the associated limitations given by the proximity link via an orbiting relay spacecraft, as direct data transfer to Earth from the surface is in most cases not possible.

10.18.4.2.6 Microscopic Imager

The Microscopic Imager is a combination of a microscope and a CCD or APS camera head that provides information on the small-scale features of rocks and soils mainly of the targets to be analyzed by Alpha Particle X-ray Spectrometer (APXS), Laser Mass Spectrometer (LMS), Raman and Mössbauer spectrometers. It reveals the fine-scale texture and resolves individual grains on soil and rocks. Such cameras usually provide multicolor images with a typical spatial resolution on the order of 50–100 μm per pixel by employing microscopic optics. Illumination devices are used when night images or images in shadow regions are required. Color information can be provided either by artificial illumination with light emitting diodes (LEDs) at different limited wavelengths or by means of wide-band illumination and filtering. Modern microscopic imagers have a typical mass for the complete system in the order of 100 g, with a power consumption of about 1 W. An example is the ROLIS imager on the Rosetta Lander (Mottola *et al.*, 2006).

10.18.4.3 Spectrometer

Spectroscopy of electromagnetic waves is a primary tool of remote sensing. It is the separation of electromagnetic waves into its component wavelengths

within the full range of the electromagnetic spectrum starting in the radio frequency (RF) band from very low frequency (VLF) with wavelength of hundreds of kilometers to extreme high frequency (EHF) with several centimeters wavelength. The microwave region covers wavelengths between 0.1 and 30 cm, higher frequencies are in the millimeter and submillimeter wavelength region. The IR and visible range covers 10^{-3} to 0.4×10^{-7} m, followed by UV range 0.4×10^{-7} down to 10^{-8} m. Extremely short wavelengths are expressed in angstroms ($1 \text{\AA} = 10^{-10} \text{ m}$). X-rays are covering 100 \AA down to 1 \AA and below 1 \AA are high-energetic gamma rays.

Certain spectral regions are quite useful to interpret specific features: radio waves for mapping gross surface properties, UV wavelengths for atmospheres and magnetospheric ions, and the near IR for understanding of mineralogical composition of solid surfaces. X- and γ -rays provide data on the surface and near-surface elemental abundances, respectively. Simply put, spectrometers measure the intensities of the individual wavelengths. Spectral information can be provided by grating, interferometric principles, filtering, heterodyne mixing, or by illumination with spectrally limited and variable emitters. The spectral information can be used to reveal chemical and elemental composition, distinguish isotopes, and characterize energy distributions or to define specific atomic energy levels. There are many applications requiring spectral analysis. Of course, spectroscopy is very often combined with imaging to provide information on spatial variations.

Due to the wide extent of the electromagnetic spectrum different detector types are required for the various bands:

1. antennas and electromagnetic coils for the RF, and microwave bands;
2. CCDs, APSs, micro-channel plates (MCPs), and bolometer for the IR, visual, and UV; and
3. various semiconductor detectors, MCPs, and scintillators for X-ray and γ -ray.

Also the optics for the various bands is very different. For example:

1. parabolic antennas (reflectors) and antenna groups are required for the higher-RF frequency bands, as well as patch antennas;
2. refractive and reflective optics in the visual, IR, and UV bands; and
3. microchannel optics for X-rays is a more recent development.

For high-resolution spectroscopy the instrument stability is a key issue. Thermal stabilization of the detector and readout is often necessary. Cooling is required for low noise applications. Calibration sources are required to eliminate instrument drifts. Sensitive detectors with very low readout noise and long observation times are required for faint objects or high-resolution spectroscopy.

10.18.4.3.1 Visual and Infrared Mapping Spectrometer

Visual and Infrared Mapping Spectrometer (VIMS) provides mineralogy and elemental composition and complement X- and γ -ray spectrometers in the visible and near-IR bands (around 0.7 to several micrometers, although improved performance can be achieved by extending the wavelength into the thermal IR up to around 60 μm). In the surface reflection spectra features occur as minima due to volume scattering bands or Christiansen features and as maxima by Reststrahlen peaks and transparency features. Infrared mapping spectroscopy (IMS) can be used to map the surface minerals and their abundances at selected locations, identify the phases and mixtures of the surface minerals, and map the mineral distributions and their correlation with images in the visible band. In addition, such an instrument can also determine the atmospheric cloud structure and its layering, the variations over time and space of the constituents of the atmosphere, and the temperature versus altitude profile, define water vapor content, as well as the abundance of dust. The instruments Visible and Infrared Thermal Imaging Spectrometer (VIRTIS) (Coradini *et al.*, 1999) and Mercury Thermal Infrared Spectrometer (MERTIS) (Benkhoff *et al.*, 2006) are examples for near- and thermal-IR imaging spectrometers.

The IR light passes a baffled instrument aperture toward an optical lens or mirror system (TMA) providing the required mapping resolution. An optical filter might be applied to reject the solar flux outside the band of interest and an entrance slit in the optical path to reduce stray light. A wave-dispersing element, which could be a beam splitter cube, a prism, a variable linear filter, a grating or a Fourier Transformation (FT) concept, is used to create and project the spectral image onto the detector array. Detectors such as InGaAs, CdTe, or HgCdTe form the focal plane assembly, which requires active thermoelectric or passive cooling. New developments in

the field of uncooled bolometers allow for operation up to room temperatures, but still however require thermal stabilized focal plane assemblies. Typical achieved instrument spectral resolutions are of the order 20 nm or less. A tiltable mirror in the optical beam is used to alternatively expose the instrument to a calibration source.

To extract reliable compositional information from IMS spectra, a thorough understanding of the spectral features from laboratory spectroscopy on analog materials is required. Of course, local ground truth which can be achieved through *in situ* instrumentation on landers and rovers is very important so as to allow for the definitive interpretation of global remotely captured spectra.

10.18.4.3.2 UV spectrometer

UV spectrometers are used to determine surface composition and structure of atmospheres, determine the loss rates of volatile gases and compositions and dynamics of exospheres, and to search for surface layers of ice. The reflective properties of natural satellite surfaces in the UV are useful to determine the composition and physical state of the materials that comprise the surface and hence complement measurements performed with IR spectroscopy and elemental compositional information obtained from X-ray and γ -ray spectrometers. The range from 70 to 800 nm contains major resonance lines of many elements and molecules such as Ne, Ar, Na, S, Mg, Al, Fe, OH, CO₂, and H₂, although it is difficult to cover the entire range with a single grating and detector.

For atmospheric limb observations the level of stray light is of major concern as it reduces the SNR of the measurement. Scanning of the different atmospheric layers during limb observation is done with a movable mechanical scanning mirror. A slit in the optical path must be used to suppress the strong background of the planetary disc. A compact instrument design can be achieved with a common Rowland spectrometer configuration. Current instrument designs use third-generation holographic gratings (Singh, 1992), which produce the UV spectra on a photon-counting MCP through a planar MgF₂ window in front of the MCP. KBr or CsI photocathodes are used as detectors for the EUV and CsTe or RbTe photocathodes for the FUV and common CCD or APS imaging array for the UV and visible range.

10.18.4.4 Radiometer and Polarimeter

Radiometers provide precise intensity data by simply establishing the brightness (or the intensity of energy radiated) from targets.

Polarimeters are optical instruments that measure the direction and extent of the polarization of light reflected from a target. They consist of a telescope fitted with a selection of polarized filters and optical detectors. Careful analysis of polarimeter data can infer information about the composition and mechanical structure of the objects reflecting the light, such as the various chemicals and aerosols in atmospheres, rings, and satellite surfaces, as these all reflect light with differing degrees and types of polarization. Of course, a polarimeter may be integrated into another instrument, such as a camera, or a photometer.

Both instrument principles require calibration sources when absolute measurements have to be provided.

10.18.4.4.1 Thermal IR radiometers

A thermal IR radiometer measures the intensity of IR energy radiated by a target, in order to determine the atmospheric and surface thermal properties and in particular the thermal inertia. When the instrument's field of view completely includes the disk of a planet, its total thermal output can be determined and as a consequence the planet's thermal energy balance derived, which in itself reveals the ratio of solar heating to the planet's internal heating, provided the latter is large enough. Thermopile arrays or bolometers are used to measure accurately parts of the Planck curve so as to establish the surface temperature. Note, the blackbody curve peak lies between several micrometers for hot bodies (e.g., Mercury) and 60–70 μm for very cold bodies (e.g., Europa). The Galileo Photopolarimeter/Radiometer Experiment (PPR)

measured thermal radiation in five spectral bands between 15 and 100 μm (Russell *et al.*, 1992).

10.18.4.5 Nephelometer

A spaceborne Nephelometer is an instrument for measuring suspended particles (aerosols) in an atmosphere to obtain information about particle properties by scattering light in a volume containing many particles (Figure 2). The instrument consists in its simplest form of a light source and a light detector, which is mounted offside the direct light beam. The aerosols in the light beam produce scattering which is measured with the detector. The light reflected into the detector is a function of the particle density, their shape, color, and reflectivity. In the simplest version only the backscattering coefficient is defined. In a more complex version the instrument measures the elements of the scattering matrix (the matrix that relates the Stokes parameters of the scattered to the incident light in the far-field limit) (Bohren, 1983). The ultimate goal is to obtain the size distribution, particle density, and refractive indices, while avoiding ambiguities caused by nonsphericity of the particles. Newer developments are based on polarimetry, as the advantages are (1) increased sensitivity to microphysical properties, (2) higher achievable accuracy, and (3) angular features insensitive to multiple scattering (Castagner and Bigio, 2006).

A light source, typically a laser operated at various wavelengths, is used with a photoelastic modulator made of a piezoelectric transducer coupled to a block of amorphous quartz to produce a periodic retardance. The incident laser light is linearly polarized and then passes through the modulator. The polarization state of the emerging light alternates between left and right circular polarization. Both the DC and the AC component of the detected scattered signal have to be measured to determine the important

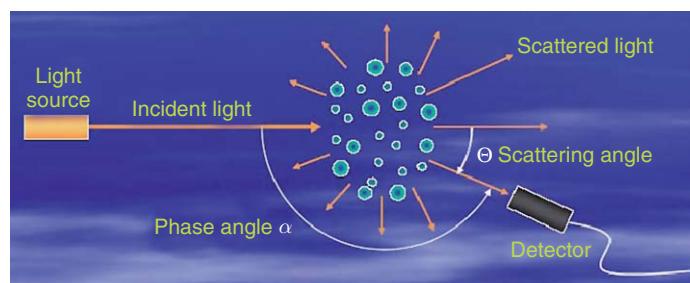


Figure 2 A nephelometer analyzes the scattering of a light source in the gas to define molecular constituents.

elements of the scattering matrix and therefore finally derive the characteristics of the suspended particles.

10.18.4.6 X-Ray Spectrometer

An X-ray spectrometer is used for remote sensing, as well as for *in situ* characterization of elemental composition through the detection of the secondary X-ray photons energy (dispersive) or photons wavelength (diffraction). Various principles are used as described in the following sections.

10.18.4.6.1 X-ray fluorescence

X-ray fluorescence (XRF) is based on the observation of (X-ray) photon emissions of atoms, which are induced into an excited state by irradiation with X-rays (e.g., originating from the Sun) causing the removal of a core electron (K, L, or M shell) from the atom. The resulting inner shell vacancies are filled by electrons from outer shells of the same atom. The difference in energy between the two electron orbitals appears as an emitted X-ray photon, which can be measured with an X-ray spectrometer either by wavelength dispersion in a diffractometer (X-ray diffraction (XRD)) or by energy dispersion with a semiconductor sensor (XRF). In both cases this leads to information on the mineralogical composition through the determination of major-, minor-, and trace-element abundances.

The XRD technique has an excellent resolution, but is mainly implemented for laboratory use, where XRF is more often used in space because of the small mass and reduced complexity. XRD instruments for space application are however currently under development.

In an imaging XRF, a micro-pore optic based on a classical Wolter-I geometry (coaxial, confocal conic sections are used to form grazing incidence imaging systems as proposed by Wolter in 1952) produces a focused X-ray spot on a detector pixel array (Bavdaz *et al.*, 2002). Semiconductor sensors based on GaAs allow for even room temperature operation.

Remote-sensing applications of the XRF technique rely on the irradiation of the target surface with X-rays for the surrounding environment (e.g., the Sun). Solar X-ray monitors (SXM) are used to observe the solar X-ray emission and to remove its spectral dependence from the XRF measurements. This is essential due to the high variability of the illuminating solar X-ray emission, and the time

variation and spectral hardening of the solar spectrum above 3 keV.

In situ instruments also use high-energy incident radiation sources such as an α -particle, proton or electron sources, and primary X-rays from the radioactive decay of selected radioisotopes such as Fe⁵⁵, Cd¹⁰⁹, or U²³⁸. This technique can therefore be used at night or in shadow regions and thus does not depend on an external source of radiation such as the Sun.

10.18.4.6.2 Particle-induced X-ray emission

Particle-induced (or proton-induced) X-ray emission (PIXE) uses bombardment with ions of sufficient energy (usually MeV protons) causing inner-shell ionization of atoms in the sample. Outer-shell electrons drop down to replace inner-shell vacancies and X-rays of a characteristic energy are emitted. An energy dispersive detector is then used to record and measure these X-rays and the intensities are then converted to elemental concentrations (e.g., Isabelle, 1994).

10.18.4.6.3 Alpha particle X-ray spectrometer

The APXS is used to determine *in situ* the chemical composition and absolute abundances of elements at the outermost few micrometers of soil samples (Rieder *et al.*, 1997). Also α -particles emitted from radioactive gases (such as radon and polonium) leaking out from a planetary interior can be detected.

The APXS has evolved from the simpler α -particle instrument (APS), which irradiates a target with α -particles emitted from an α -source such as Cm²⁴². The energy distribution of the backscattered α -particles is measured to identify light elements (except hydrogen) including rock-forming elements like Na, Mg, Al, and Si. When a backscattered α -particle hits the silicon wafer detector, a small track of charge is created, which is sensed with a charge sensitive amplifier (CSA). The charge produced is directly proportional to the energy of the backscattered alpha particle. By measuring integrated count rates over energy distribution the elemental composition of the sample can be defined.

An improved version of the instrument, the APXS, also analyzes the X-rays produced in the sample. A Cm²⁴⁴ source with a typical activity of around 50 mCi emits energetic α -particles and X-rays, which are used to irradiate the surface of the sample area at a very close distance. Hard X-rays induce

fluorescence of elements with high atomic number. Medium atomic number elements, such as calcium, are excited by α -particles and X-rays. All emitted X-rays are measured with a small solid-state silicon detector over an extended period of time and a spectrum is determined. The energy peaks in the spectrum characterize the elemental composition. The intensity in the peaks reveals the concentration of each element. The APXS is sensitive to major elements, such as Na, Mg, Al, Si, K, Ca, and Fe, and minor elements, such as P, S, Cl, Ti, Cr, Mn, and Ni. Due to the short ranges of α -particles and X-rays, the sampling depth is less than 10 μm . The APXS typically requires long integration times of order 1–2 h per sample to achieve sufficient counting statistics. Longer measurements and cold environment (night measurements) improve the counting statistics of peaks of lower intensity from elements having a lower abundance.

The APXS is preferably mounted on a robotic arm or a rover to provide mobility and access different locations for chemical compositional analysis.

10.18.4.7 Gamma-Ray Spectrometer

Gamma-ray spectroscopy (GRS) is a well-established technique for determining the elemental composition of the surface of planetary bodies. Measurements can be performed either from orbit or directly on planetary surfaces. *In situ* measurements on surfaces are important since they provide direct analysis without the need to compensate for atmospheric effects or other contributions. Thus, surface measurements can verify and extend the interpretation of orbital measurements and provide a better estimate of the elemental composition, but are in most cases limited to local measurements only.

The detected gamma rays arise from two sources: gamma rays emitted spontaneously by naturally occurring radioactive elements like K, Th, and U and cosmic-ray induced gamma rays emitted by elements like H, C, O, Si, and Fe. The nucleus of each chemical element produces a unique set of gamma-ray lines, and the technique of gamma spectroscopy allows the unique identification of these lines and therefore the specific chemical element and its relative abundance. The ambient cosmic-ray flux produces neutron-induced reactions on elements in the planetary surface, which in turn produce the characteristic γ -rays that are used to determine the elemental concentration. A scintillator (e.g., cesium

iodide (CsI), LuAP, or LaBr₃) is used to convert the γ -ray into light emission, which is detected by sensitive large photodiodes. The background provided by cosmic rays and by the spacecraft is a source of noise. Mars Odyssey GRS, for example, was mounted on a 6 m boom to reduce the S/C background for the measurements (Boynton *et al.*, 2004). The use of anti-Compton shielding, of the same or different detector material, is often used to improve the quality of the received spectra by rejection of γ -rays that do not come from the planetary surface, or are resulting from cosmic-ray interactions with the spacecraft material. Important instrument factors are the relative precision of the scintillator material, sensitivity to radiation damage, quantum efficiency of the readout photo diodes, and the capability to suppress background counts. Since the gamma-ray spectrometer sensitivity increases with the square root of the integration time, the expected integration times range from several orbits for the natural radio nuclides to several months for the cosmic-ray induced events. Of course, the sensitivity is also a strong function of the efficiency and energy resolution of the instrument.

Surface compositions finally are deduced from a comparison between the γ -ray spectra emitted by an area of the planetary surface with simulations based on calculated fluxes of assumed surface composition. By applying an iterative process, it is possible to derive a chemical composition that best fits the measured in-orbit fluxes.

Examples of gamma-ray spectrometers based on either Ge-detectors or scintillators are the gamma-ray and neutron spectrometer (GNRS) on Messenger (Burks *et al.*, 2004) or the Mercury Gamma-ray and Neutron Spectrometer (MGNS) (Kozyrev *et al.*, 2006) on the BepiColombo Mission to Mercury.

10.18.4.8 Neutron Spectrometer

A neutron spectrometer (e.g., the HEND instrument on Mars Odyssey (Mitrofanov *et al.*, 2002) is designed to detect neutrons in three energy bands: thermal (average energy of about 0.025 eV), epithermal (between 0.025 and 1 eV), and fast neutrons. Each energy class corresponds to the degree to which planetary neutrons have been ‘moderated’, or been in contact with other planetary matter. These free neutrons are produced in collisions between galactic cosmic rays and planetary matter. Hydrogen is a very good moderator of neutrons and hence a neutron spectrometer is quite sensitive to the presence of

hydrogen on the surface (to a depth of about 1 m). Large concentrations of hydrogen are most likely indicative of water in liquid or ice form. Hence, the main purpose of a neutron spectrometer is to search for, and determine the abundance of hydrogen in form of buried water ice. The neutron flux and its energy dependence is a complicated function of the chemical composition of the surface and especially of its moderation by light elements, such as hydrogen and carbon.

The neutron detector is, for example, a neutron-absorbing boron-loaded plastic scintillator (e.g., BC454, GS20 glass). Neutrons lose energy in the detector through multiple scattering collisions with the hydrogen and carbon nuclei that compose the scintillator. Multiple ion-electron pairs are produced during that process that eventually recombine to produce photons. The light emission is detected with a photomultiplier tube (PMT) or large sensitive photo diode with a high quantum efficiency.

Thick sheets of cadmium can be used to shield parts of the detector to achieve some crude spatial information or to make the detector only responsive to neutrons having energies above a certain range. It is important to distinguish neutrons coming directly from the planetary surface and the background neutron flux. This can be achieved by using separate detectors facing different directions in combination with appropriate types of shielding.

A detector facing in the direction of the spacecraft velocity vector encounters thermal neutrons, which travel slower than the spacecraft. An additional backward detector detects thermal neutrons coming from the planetary surface. Using the combination of geometry and velocity, a separation of background and spacecraft created neutrons from those emitted by the planetary surface is possible. The difference in counting rates between the forward and backward detectors yield a measure of the flux of thermal neutrons.

Differentiation between the different energy bands (thermal, epithermal, and fast) of the neutrons can be established through the use of different scintillator types in combination with shielding and moderation.

Readout of the instrument is achieved by means of a PMT or photodiodes operated with coincidence and anticoincidence logic to separate different energy bands and flux directions. The energy of these neutrons can be measured by determining the amplitude of the scintillator's light flash.

Finally, the instrument data outputs are typically histograms of the measured amplitudes integrated over time.

10.18.4.9 Laser Altimeter

A classical laser altimeter (also called Light Detection and Ranging or Laser Image Detection and Ranging (LIDAR)) transmits laser pulses (energy in the order of 50 mJ per pulse) toward planetary surfaces with a typical pulse rate of 10–30 Hz and measures the time of flight (TOF) of the light pulse to determine the range of the spacecraft with respect to the surface. A laser altimeter operates in a very similar way to radar technology, which uses radio frequency pulses instead of light pulses. The laser pulses are produced, for example, with a neodymium-doped yttrium aluminum garnet (Nd:YAG), which is optically pumped with laser pumping diodes. Active or passive Q-switching is used to produce the controlled laser pulses. The laser pulse is typically in the IR for an Nd:YAG (mainly 1064 nm), but with other YAG doping also light emission in UV, visible, or near IR can be produced. A laser typically has a very narrow beam which allows the mapping of surface features with very high spatial resolution compared with radar.

An optical telescope is used to focus the reflected laser pulse onto an Avalanche Photo Diode (APD), which is a high biased Si-photo diode, which is sensitive to the received photons and creates a charge avalanche which is detected with a CSA. A precise time measurement between pulse emission and return signal detection is important for the precise functionality of a laser altimeter.

Repeated range measurements over consecutive orbits are used to constrain the morphology of planetary surfaces, construct precise topographic maps and digital elevation models (DEMs). Repeated measurements over the same area at various seasonal periods can be used to study the stress and strain and flexural properties of the lithosphere, determine tectonic movements and tidal deformations, and can be combined with gravity data to study the density distribution in the crust.

A LIDAR is highly sensitive to aerosols and cloud particles and thus has also many applications in atmospheric research and meteorology.

The receiver of a laser altimeter can be also used as a passive radiometer measuring the radiance of the surface at its sensitive wavelength. To increase the signal-over-noise and reduce the false measurement

rate very narrow filter (only several nanometers pass window at the laser emission wavelength) is used in the receiver optical path. The MLA altimeter flying on Messenger and BELA on BepiColombo ([Spohn et al., 2005](#)) are examples of recent laser altimeter instruments.

Emerging single photon counting laser altimeters emit laser pulses at much higher repetition rates in the order of 20 kHz, but at much lower energy (typically 20–50 µJ). Due to the low photon counts of the return signal a statistical analysis of many pulses needs to be performed to retrieve the range measurement. Planetary single photon counting laser altimeters are currently under development and promise significant reduction in mass and power.

10.18.4.10 Radar

Several solar system objects are covered by clouds or haze, making optical imaging difficult or impossible. But their atmospheres are transparent to RF waves and hence use can be made of radio detection and ranging (radar) instruments. Radars are active instruments, which illuminate a target with microwaves allowing detailed observations at any time, regardless of weather conditions, opaque atmospheres (in UV, IR, and visual band) or presence of sunlight illumination.

In principle, the full range of the radio waveband ([Table 1](#)) from kilometer to submillimeter wavelength can be used, according to the application and the particular requirements. The selected frequency depends on considerations, such as antenna size, required footprint size, thermal background noise, wave interaction with the target surface and wave interaction along the transmission path and the interaction with a possible ionosphere.

10.18.4.10.1 Radar altimeter

Radar altimeters send either radio pulses (pulsed radar) or continuous wave signals (CW-radar) straight down to a planet's surface (the nadir) to measure variations in the height of the terrain below. In the case of the pulsed radar the signals are timed from the instant they leave the instrument until they are reflected back. From the round-trip time the distance can be obtained. Actual terrain height is then deduced based upon precise knowledge of the spacecraft's orbit – derived by laser or radio ranging.

CW radar altimeter often uses frequency modulation (FM) with a relative slow linear ramp function (for practical reasons a triangular frequency

modulation is used). During the round-trip time of the emitted signal, the local oscillator increased its frequency because of the applied ramp modulation. This causes a frequency difference between the transmitted and received signal. The longer the round-trip time, the larger the frequency difference becomes. Hence, the measurement of the frequency difference determines the range to the target (see [Skolnik \(1980 pp. 70–92\)](#)).

Laser altimeters generally have a smaller footprint, and thus higher spatial resolution, than radar altimeters. They also require less power. Nevertheless there are many applications where the use of radar altimetry is beneficial, particularly in Earth observations ([Gommenginger et al., 2006](#)).

10.18.4.10.2 Ground penetrating radar

For ground penetrating radar (GPR) relative low frequencies in the megahertz range are selected, because signal penetration depth primarily decreases with increasing frequency. Such low frequencies require long antenna with several tens of meters length. For instance, the MARSIS instrument on Mars Express has 40-m-long antenna ([Biccari et al., 2002](#)). At low frequency however, two other fundamental aspects of the GPR measurement come into play: lower frequency results in a loss of spatial resolution and if the frequency becomes too low, electromagnetic fields can no longer travel as waves but with frequencies in the megahertz range ground penetration in dry soil down to several kilometers can be achieved. A large part of the signal intensity is reflected at the surface, but a significant fraction of the signal travels through the planetary crust and is reflected at subsurface interfaces between layers of different material, including water or ice. Water is a strong attenuator due to its high permittivity and hence subsurface layers of liquid water can be relatively easily detected by GPRs. Nevertheless, the surface reflection and surface clutter make the GPR data analysis difficult.

10.18.4.10.3 Synthetic aperture radar

Synthetic aperture radar (SAR) synthesizes the angular resolving power of an antenna many times the size of the antenna aperture actually used. SAR illuminates its target to the side of its direction of movement, and travels a distance in orbit while the reflected, phase-shift-coded pulses are returning and collected. The SAR emits a permanent series of pulses as it travels and the amplitude and phase of the individual returning signals are recorded.

Table 1 Electromagnetic wave-band definition and radio spectrum nomenclature

Radio spectrum nomenclature										
ELF	SLF	ULF	VLF	LF	MF	HF	VHF	UHF	SHF	EHF
3 Hz	30 Hz	300 Hz	3 kHz	30 kHz	300 kHz	3 MHz	30 MHz	300 MHz	3 GHz	30 GHz
30 Hz	300 Hz	3 kHz	30 kHz	300 kHz	3 MHz	30 MHz	300 MHz	3 GHz	30 GHz	300 GHz
<i>Band name</i>		<i>Frequency range</i>			<i>Wavelength range</i>			<i>Instrument application</i>		
HF		3–30 MHz			10–100 m			Ground penetrating Radar		
P		<300 MHz			1 m+					
VHF		50–330 MHz			0.9–6 m					
UHF		300–1000 MHz			0.3–1 m					
L		1–2 GHz			15–30 cm					
S		2–4 GHz			7.5–15 cm					
C		4–8 GHz			3.75–7.5 cm					
X		8–12 GHz			2.5–3.75 cm					
K _u		12–18 GHz			1.67–2.5 cm					
K		18–27 GHz			1.11–1.67 cm			Water vapor measurements (22 GHz)		
K _a		27–40 GHz			0.75–1.11 cm					
Mm		40–300 GHz			1–7.5 mm			Millimeter wave cloud radar		
V		40–75 GHz			4.0–7.5 mm					
W		75–110 GHz			2.7–4.0 mm			Millimeter and submillimeter sounder		

A single pulse from the antenna is still rather broad (several degrees) because diffraction would require too large an antenna to produce a narrower beam. The results from many pulses can be combined using extensive onboard computer processing to create a synthetic aperture much larger than the length of the actual antenna used. The spacecraft's position and velocity must be known with great precision, and its attitude must be controlled tightly for the complex data processing, based on FT techniques (see Skolnik (1980 pp. 517–529)). The required complex data processing on board the spacecraft and the high produced data volume make the use of SAR for planetary mission difficult and hence SAR is mainly limited to Earth observations.

10.18.4.10.3.(i) Interferometric SAR Interferometric SAR uses the phase information of the returned pulses. Any phase difference of the return signals obtained simultaneously by two spatially separated antennas on the spacecraft contain information about the angle from which the radar echo returned. Combining this with the distance information, the position in three dimensions of the image pixel can be derived.

If accommodation of two receiver antennas is not possible, measurements separated in time, perhaps from two different orbits, can be used to derive the information based on the phase shift, which has occurred similar to that described above. Even a terrain shift between observations (e.g., due to tides) is represented in the return phase difference and hence can be detected, provided the overall system resolution is sufficient and proper orbit and attitude data is available.

10.18.4.10.3.(ii) Polarimetric SAR Anisotropic materials reflect different polarizations with different intensities. Some materials also convert one polarization into another. By emitting a mixture of polarizations and using receiving antennas with a specific polarization, several different images can be collected from the same series of pulses. Such images can be used as kind of color information in a synthesized image. Interpretation of the resulting colors requires significant testing of known materials.

10.18.4.11 Radio Science

Radio-science experiments are performed to determine masses and internal structures of bodies from spacecraft tracking data and to determine satellite

radii and atmospheric structure from radio propagation information.

10.18.4.11.1 Celestial mechanics and the gravitational field

The spacecraft communication radio subsystem can also be used to sense small changes in the trajectory of the spacecraft due to gravitational field variations. For range-rate (or velocity) measurements the spacecraft's radio transmitter sends a signal at a well-known stable frequency, generated by an ultra-stable oscillator (USO) onboard the spacecraft. Any change in speed that the spacecraft experiences will cause a frequency change of the radio signal received on Earth due to the Doppler effect. The spacecraft is pulled by the gravity when it passes close to a planet, moon or any other massive body, causing its speed to change. The introduced change depends on the spacecraft distance from the body, on the total mass of the body, and the internal mass distribution. Thus, by measuring the change in frequency (Doppler) of the Earth-received radio signal, the mass and internal structure of planetary bodies can be assessed. Only velocities in the direction of the wave propagation (along the ground-station to satellite line) cause a frequency change; hence, observations are limited to spacecraft velocity changes along that direction. The closer a spacecraft approaches the body, the stronger it encounters the gravitational effects. But atmospheric drag or spacecraft internal perturbations limit the achievable resolution. A three-axis high-sensitivity accelerometer onboard the spacecraft disentangles the gravitational from other forces, even allowing for the detection of local subsurface mass concentrations (mascons).

By precise measurement of the transmission time from the ground station to the spacecraft and back (the spacecraft just ‘echoes’ the received signal), an absolute distance for ground station to spacecraft can be derived, called a range (or distance) measurement. Plasma effects and any other propagation distortions in atmospheres (like rain, water content, or change in the total electron concentration) along the line of transmission influence the measurements. A simultaneous dual-frequency link (e.g., X- and Ka-band) can be used to minimize such effects and increase the resolution (Less and Boscagli, 2001).

10.18.4.11.2 Radio propagation at occultation

The propagation of the spacecraft radio signal can be used to investigate planetary atmospheres and ionospheres. This is performed with radio

occultation experiments, when the spacecraft passes behind a planet or satellite as viewed from Earth.

The radio signal propagating from the spacecraft to Earth experiences both refraction and scattering in the atmosphere of the occulting body. The atmosphere will bend and slow the radio signal by the process of refraction; additionally, the atmosphere will diffuse the electromagnetic waves of the signal by the scattering process. This causes changes in the frequency and amplitude of the signal received at the ground station on Earth. Analysis of these changes yields information about the atmospheres and ionospheres. Unfortunately, the signal has to pass (in most cases unless there is a satellite-to-satellite radio experiment) also the ionosphere and atmosphere of the Earth and is influenced in the same way as described above. However, the influence of the Earth's atmosphere can be estimated by means of additional measurements (radio experiments, water content definition, total electron content measurements to mention just a few) and by the location of ground stations at positions on the Earth where the influence is minimal.

10.18.4.12 Microwave Sounder

Microwave sounders are powerful remote-sensing instruments complementing measurements in the IR-band with those in the (sub) millimeter band. A microwave sounder can be configured both as a continuum radiometer or a very high spectral resolution line receiver to measure naturally occurring microwave thermal emission from planetary surfaces or atmospheres, or to remotely sense vertical profiles of atmospheric gases, temperature, pressure, and cloud ice (Gulkis *et al.*, 2006). Various frequencies are analyzed to derive atmospheric vertical temperature profiles or to determine the water vapor content (clouds, precipitation, and humidity) at various altitudes. Microwave observations are particularly sensitive to the reduced forms of oxygen and nitrogen (water and ammonia), using centimeter-to-decameter wavelength to analyze, for example, abundances in the deep, well-mixed portion of the Jovian troposphere. Data are collected globally to determine horizontal and vertical variations in abundances, which can be used to investigate general circulation of the atmospheres and the nature of convective and cloud-forming processes. Microwave sounding is therefore a uniquely suited remote-sensing technique to address key questions

about (giant) planetary atmospheres (Huestis *et al.*, 2002).

The benefit of a passive microwave sounder is not only given by the global coverage, compared to local measurements of (multi) probe *in situ* missions like that of the Galileo Probe, but also are significantly less challenging than entry probes.

A single parabolic antenna (or an antenna group used mainly in Earth observation due to increased mass) is used to receive and focus the natural microwave emission onto the (very) low-noise amplifier (LNA). Depending on the application either several very narrow bands, for particular molecular abundance characterization, or a wider band for thermal emission characterization are used.

10.18.4.13 Particle Detector

10.18.4.13.1 Heavy ion counter

A heavy ion counter (HIC) instrument monitors very high energy heavy ions hitting the detector. Heavy ions observed during solar flares have, for example, been analyzed to determine the composition of the Sun. Measurements are also of engineering interest as they can cause random changes in a spacecraft's electronics. These heavy ions are detected by using stacks of single-crystal silicon wafers. Such a HIC can measure heavy ions with energies from a few up to hundreds of megaelectron volts per nucleon. A good example is the HIC on the Galileo Spacecraft (Garrard *et al.*, 1992).

10.18.4.13.2 Energetic particle analyzer

An energetic particle analyzer (EPA) uses silicon solid-state detectors and a TOF detector system to measure the energy and angular distribution, composition, and stability of trapped radiation within a planetary environment. They are also essential for the study of the interaction of energetic particles with natural satellites and the solar wind so as to derive thermal plasma flow velocities and their temperatures and to examine adiabatic and nonthermal processes in the trapped radiation. They measure the numbers and energies of ions and electrons whose energies exceed about 20 keV, define the direction of travel of such particles, and can determine the ion composition. For more information see Funsten *et al.* (2004).

10.18.4.13.3 Neutral particle analyzer

A neutral particle analyzer (NPA) needs to discriminate charged particles from neutrals. This is done by applying a high voltage (in the order of a few

kilovolts) across collimating optics, which repels charged particles. By switching off this voltage the instrument is also able to measure charged particle (super thermal ions and electrons). The mass analysis of the neutrals is based on a TOF principle. An ultrasonic chopper located at the entrance point of the instrument provides a time masking, which allows neutrals to enter only at a definite and very short time interval. The time difference between passing the chopper and the impact time on the detector MCP or solid-state detector (SSD) is used to define the mass of the neutrals, which is given by the impact energy and the measured TOF. Detectors can be designed to have a spatial resolution (pixilated SSD or MCP with sector readout), which allows not only for the TOF but also the definition of the trajectory of the incoming particles. NPAs are used to measure, for example, the escaping neutral gas composition and dynamics of planetary atmospheres and exospheres and their interaction with the solar wind, solar radiation, and meteor bombardment. The Rosina instrument on board Rosetta has unprecedented capabilities in determining the composition, velocities, and temperatures of both, neutrons and ions (Balsiger *et al.*, 1998).

10.18.4.14 Mass Spectrometer

Mass spectrometry is useful to characterize atoms or molecules in a sample, as well as determining chemical and structural information about molecules. Mass spectrometers use the difference in mass-to-charge ratio of ionized atoms or molecules to separate them from each other. All commonly used mass analyzers use electric and magnetic fields to apply a force on charged particles (ions).

The general operational steps for a mass spectrometer are:

1. create gas-phase ions,
2. separate the ions in space or time based on their mass-to-charge ratio, and
3. measure the quantity of ions of each mass-to-charge ratio.

The different principles for mass spectrometer in use are:

1. Magnetic focusing analyzer.
2. Quadrupole mass analyzer.
3. Ion trap.
4. Wien filter.
5. TOF mass analyzer.

10.18.4.14.1 Magnetic focusing analyzer

Magnetic sector-field mass spectrometers typically consist of an ion source, a magnetic momentum analyzer, and channeltron electron multipliers or MCPs as detector. Ions, formed by electron bombardment are accelerated and focused by an applied electrical field (in the ion optics) and collimated into a beam, which then traverses a magnetic field that bends it onto a collector slit. The mass-to-charge ratio of ions that successfully traverse the curved path of the mass spectrometer and pass the slit is a function of the applied magnetic field strength and the ion momentum. This can be varied to select different ion masses. Some spectrometers use multiple collectors, with multiple different path radii focused simultaneously on multiple collector slits, to cover a wide mass range but with a narrow voltage sweep range. In this way, multiple mass peaks may be monitored simultaneously.

Behind each slit high-counting-rate channeltron electron multipliers monitor the ion beams by counting the arriving ions. A specific ion mass can be selected with a particular voltage applied to the ion optics. Scanning through a voltage range provides a counting rate, which is a function of the applied voltage and allows for the determination of a mass spectrum as the resulting output.

Magnetic mass spectrometers may be combined with electrostatic analyzer to achieve double-focusing for both bending angle and energy which leads to a higher mass resolution than single-focusing magnetic analyzers.

10.18.4.14.2 Quadrupole mass analyzer

The conventional quadrupole mass analyzer utilizes four parallel cylindrical or hyperbolically shaped rods (**Figure 3**). The rods are long relative to the inner surrounding diameter, to minimize fringing fields on the active length of the rods. A quadrupole potential is established by applying a time-varying potential on alternate rods with an RF (up to several megahertz frequency) and a DC component.

The ions are injected into the rod arrangement and only those ions having the correct mass-to-charge ratio are transmitted to the exit aperture without colliding into one of the rods. Hence the quadrupole arrangement works as a mass filter. A full mass spectrum is obtained by sweeping the DC offset and the RF amplitude at a fixed frequency, and detecting the transmitted masses over the sweep at the exit plane with a Faraday cup or particle multiplier. The resolution of the device depends on the

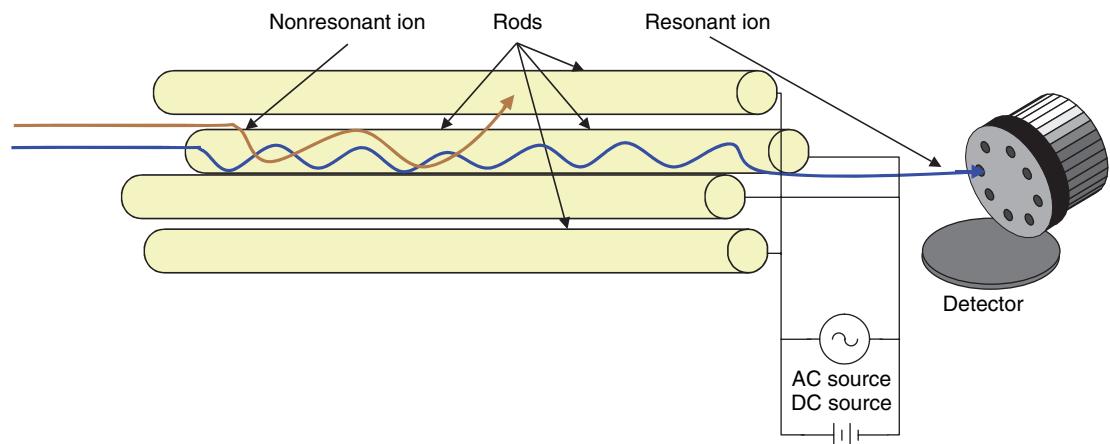


Figure 3 Principle of a quadrupole mass spectrometer.

rod geometry, frequency, rod length, and axial and radial ion injection energies.

Improved quadrupole mass analyzers make use of hyperbolic rods with narrow wings, resulting in a better mass resolution. A miniature quadrupole array has already been constructed with a mass of several grams (excluding control and readout electronics). The future challenge is to miniaturize the control and readout electronics, as well as the ionization source.

10.18.4.14.3 Ion trap

A quadrupole ion trap uses three electrodes to trap and accumulate ions in a small volume (Figure 4). A middle ring electrode separates two hemispherical electrodes that are connected to ground potential. The middle ring is connected to an RF voltage with frequencies up to several megahertz with some DC offset. Ions brought into this arrangement are trapped in the RF field and perform stable oscillations between the electrodes. Ions of a particular mass-to-charge ratio can be ejected by increasing the RF signal amplitude. A full mass spectrum is obtained by sweeping the electrode voltages over the full range to eject all ions having different mass-to-charge ratio from the trap into an electron multiplier. A full spectrum can be obtained by counting the events at the electron multiplier output as function of the electrode control voltage. Ion traps have a very excellent mass resolution with good signal-to-noise behavior since the trap effectively works as an accumulator of input ions. Full ion trap analyzers can be designed with a mass of less than 100 g.

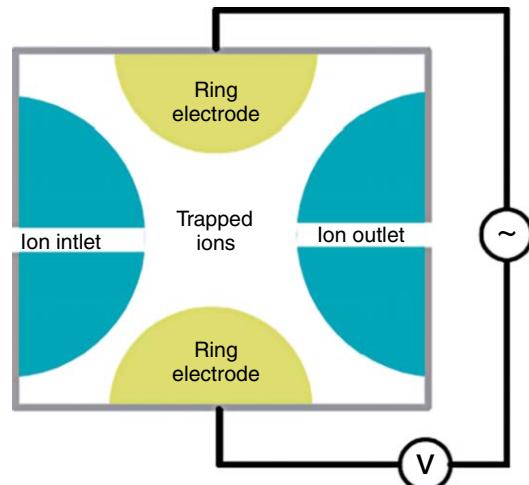


Figure 4 Principle of an ion trap.

10.18.4.14.4 Wien filter

The Wien filter is based on the principle where ions traverse a region with a crossed electric and magnetic field on an injection trajectory perpendicular to the crossed fields. Only ions with the right injection energy will pass this region linearly, when the electrical force (Coulomb force dependent on ion charge and electric field strength) just matches exactly the magnetic deflection (Lorentz force dependent on ion velocity, ion charge, and magnetic field strength), provided that a good homogeneity and alignment of the magnetic and electric field is given. All other ions with different injection energy will be deflected from the linear trajectory as the electrical force does not match the magnetic force. A spatially extended detector (MCP, CCD, or multi-anode detector)

after the Wien filter is used to detect different deflected ions and to create a mass spectrum. A Wien filter can be further used to remove ions of unwanted isotopes and impurities, multiple charged ions, and neutral particles.

10.18.4.14.5 Gas chromatograph mass spectrometer

Gas chromatograph mass spectroscopy (GCMS) is used to measure abundances and isotopic ratios of noble gases He, Ne, Ar, Kr, and Xe and the abundances of minor gas constituents like SO₂, COS, CO, HCl, H₂S, and H₂O (non-exhaustive) in planetary atmospheres. A low gas-flow rate at a reduced pressure is produced with a pump system controlled with valves to obtain fresh gas samples combined with a carrier gas (such as hydrogen) injected into the gas chromatograph (GC). The GC provides a time-dependant separation of the gas components prior to a molecular identification in a mass spectrometer (MS). The output of the GC section is ionized with, for example, electron impact ionization before arriving in a quadrupole mass filter, followed by an ion detector.

Gas chromatograph mass spectrometer systems are among the most powerful analytical tools for chemical analysis of many compounds and especially of gas mixtures, with however the drawback of having an increased instrument complexity. GCMS instruments have been flown for example on Huygens ([Niemann et al., 2002](#)) and the Rosetta Lander Philae ([Goesmann et al., 2006; Wright et al., 2006](#)).

10.18.4.14.6 TOF mass analyzer

A TOF mass spectrometer consists of an ion source, accelerator, timing gates, a long field-free flight tube, and particle detector. An electric field is used to accelerate the source ions into the drift region of the TOF. Gates at the entrance of the drift tube are opened only for a short time to allow a group of ions to enter the drift tube. Ions having a different mass-to-charge ratio take different transit times through the drift region. Lighter ions have a higher velocity than heavier ions and hence reach the detector at the end of the drift region earlier. To avoid mass-aliasing effects additional gates are introduced in the drift region, before the detector and at the end of the drift region to eliminate aliasing velocities. A TOF analyzer has a relatively large mass range and a good sensitivity provided the drift tube is not too small.

10.18.4.14.7 Laser ablation mass spectrometer

Here a short well-focused pulsed laser beam is used to ablate small amounts of the surface material from a target and converts it into low-energy plasma. The ablation can be done even without prior sample preparation. Resulting ions are analyzed with a TOF mass spectrometer, which determines elemental and isotopic composition of the ablated material. The electrostatic fields in the reflectron of the TOF are designed to cancel energy dispersion of the ions over a broad energy range covering the peak of the ion distribution. This results in the ion TOF from the point of ablation to the MCP that is independent of ion energy and proportional to the square root of ion mass. For typical compact laboratory instruments, the TOF ranges from a few microseconds (for H) to <100 µs (for >200 amu). The MCP current as a function of time relative to the laser pulse provides a mass spectrum of the analyzed material. The MCP output current can be digitized with high accuracy and high time resolution, thus providing isotopic mass resolution for various elements. A laser ablation mass spectrometer was flown on the Phobos Mission, LIMA-D (e.g., [Pellinen et al., 1990](#)). Modern instruments are small, low power, and low mass (250 g instruments are under development). These are ideally suited for missions to planets and small bodies in the solar system to provide quantitative analysis for major elements and isotopes present in a sample down to levels below 100 ppm. See [Brinkhoff \(2005\)](#) for an analysis of the application to small bodies.

10.18.4.14.8 Laser-induced breakdown spectroscopy

Laser-induced breakdown spectroscopy (LIBS) or Laser-induced plasma spectroscopy (LIPS) is atomic-emission spectroscopy based on a high-energetic laser pulse as the excitation source. It consists of a solid-state laser and a spectrometer. The laser emits short (typically 10 ns) focused pulses with a local power density of around 10^7 W m^{-2} at the very small focal point on the sample. This high energy ablates small amounts of material and converts it into a plasma plume which breaks down into excited ionic and atomic species. At these high temperatures the plasma cannot be analyzed as it emits continuum radiation, but the plasma rapidly expands and cools (within 10 µs) and consequently reveals the characteristic atomic-emission lines of the elements contained, which is analyzed with a spectrometer with a polychromator (Echelle type) and a CCD or

APS detector for readout. The spectrometer response is typically from $1.1\text{ }\mu\text{m}$ down to 170 nm covering almost all elements and within the sensitivity range of a classical CCD or APS detector.

Major advantages of the LIBS technique are its ability to provide a depth profile of a sample by repeatedly focusing the laser pulses on the same spot and analyzing the generated plasma. Spectra can be produced within seconds (compared to hours with APX or Mössbauer). LIBS also allows for some remote analysis of samples by employing telescopic optics. For proper measurements the laser output needs to be kept stable.

A LIBS utilizes many components that are also used for a Raman spectrometer. A combined LIBS–Raman spectrometer allows for both atomic and molecular characterization of a sample. The LIBS–Raman principle is described in Giakoumaki *et al.* (2006). Currently combined miniaturized LIBS–Raman instruments for space applications are under development.

10.18.4.14.9 Raman spectrometer

Raman spectrometer uses the inelastic (Raman-) scattering of light to analyze vibrational and rotational modes of molecules. Monochromatic light, produced from a laser in the visible, near-IR, or near-UV range is scattered by a molecule or by a crystal. One photon of the incident radiation is annihilated and, at the same time, one scattered photon is created. In the case of elastic Rayleigh scattering the energy of the scattered photon is equal to that of the incident one. Alternatively the energy of the scattered photon is different to that of the incident one, when inelastic Raman scattering occurs. Spontaneous Raman scattering is typically very weak, and as a result the main difficulty of Raman spectroscopy is separating the weak inelastic scattered light from the intense Rayleigh-scattered laser light.

The spectrum of the scattered light consists of a strong exciting line of the same wavelength as the incident illumination due to the Rayleigh scattering together with weaker Raman lines at different wavelengths. The individual shift in energy gives information about the phonon modes in the illuminated matter. The lines having longer wavelength are called Stokes lines, the others anti-Stokes lines.

In a Raman spectrometer the sample has to be illuminated with a laser beam. The light from the illuminated spot is collected with a lens and sent through a monochromator. Wavelengths close to the laser emission line are filtered out (to remove

the elastic Rayleigh scattering) and those in a certain spectral window away from the laser line are dispersed onto a detector. Typically holographic diffraction gratings and multiple dispersion stages are used to achieve a high degree of laser rejection. A photon-counting PMT or a CCD can be used to detect the Raman scattered light.

10.18.4.15 Mössbauer Spectrometer

A Mössbauer spectrometer determines *in situ* the mineralogical composition of soil samples and provides quantitative analysis of Fe-bearing materials. It can determine the composition and abundance of these minerals, as well as their magnetic properties to a high level of accuracy. It makes use of the resonance absorption of γ -rays by Fe^{57} nuclei due to the Mössbauer effect. The absorption provides information on the nuclear energy level splitting of Fe-bearing materials. The splitting reflects the nuclear levels occupied by surrounding electrons.

Co^{57} is normally used as the γ -ray source. In general the structure of the nuclear energy levels of the soil samples differ from that of the Co^{57} γ -ray source because of different oxidation states, chemical environment, and magnetic order. To achieve resonance with the soil samples the energy levels of the emitted γ -rays from the source must be modulated to overlap with energy levels of the sample. The Co^{57} source is mounted on an electromechanical velocity transducer and moved with respect to the target, leading to velocity-modulated energy levels of the emitted γ -rays due to the Doppler effect. The energy increases if the source moves toward the sample and decreases when moved away. Phase and the oxidation state of the target can be determined from the peak locations in the resulting spectrum. The concentration can be estimated from the peak intensities in this spectrum.

Analysis of material can be performed by either transmission or backscattering, although the transmission method is limited to thin targets directly placed between the source and the detector.

With the backscattering principle only the surface of the sample contributes to the received spectrum. The average diagnostic depth for a backscattering Mössbauer spectrometer is in the order of 100–200 μm (assuming basaltic rock composition). The backscattering approach is generally preferred for planetary applications.

The Mössbauer spectrometer is sensitive to temperature variations, which should not exceed about

$\pm 10^{\circ}\text{C}$ during acquisition of a single spectrum to minimize the temperature-dependent parameter variations and the resulting spectral smearing. Measurements are taken by placing the instrument's sensor head directly against a rock or soil sample. A simple Mössbauer measurement typically can take about 12 h. Mössbauer spectrometers are preferably mounted on micro-rovers or robotic arms to provide mobility (e.g., Wade *et al.*, 1999).

10.18.4.16 Dust Analyzer

A dust detector instrument is an *in situ* instrument used to measure the mass, electrical charge, and velocity of incoming dust particles. The masses of dust particles which can be detected are typically in the range $10^{-16}\text{--}10^{-6}\text{ g}$. The speed of these small dust particles is measured over a typical range of 1 km to several tens of kilometers per second at impact rates from 1 particle per 100 days up to several hundreds of particle impacts per second. The instrument aperture is the important characteristic to guarantee a proper count rate for a specific application. The length of the flight path of the TOF mass analysis section defines the achievable mass resolution and contributes, together with the aperture, to the overall instrument size. Dust compositional analysis instruments have flown on various comet missions (e.g., Kissel *et al.*, 2003).

A quartz crystal microbalance (QCM) can also detect tiny mass deposits on the sensor crystal by the resonance frequency change introduced by the addition of the small mass. Frequency measurements can be made with high precision; hence, it is easy to measure small mass increments with a QCM. Correlation between mass and frequency is achieved by means of the Sauerbrey equation (Sauerbrey, 1959). For compensation of the strong temperature dependence a sealed reference quartz system is required, which is exposed to same temperature, but is protected from additional particle deposits. An example of a QCM flown in space is that forming part of the Electric Propulsion Diagnostic Package (EPDP) onboard SMART-1.

10.18.4.17 Radiation Environment Monitor (REM)

To monitor the radiation during cruise phase and operational phase is not only important from an engineering point of view, but also scientifically of interest. Observations of solar flares or measurements

of the Jovian radiation belts are some examples. A radiation monitor has to cover a wide energy range starting at several kiloelectronvolts up to gigaelectronvolts and must be sensitive to electrons, protons, heavy ions, and galactic cosmic rays. Only a combination of various detectors (SSDs, scintillator, and scintillating fiber) can satisfy these requirements. Important issues are: (1) the flux measurement to derive the energy spectrum and (2) the measurement of the total (integrated) dose.

Newer technologies employ thermoluminescent dosimeters (TLDs) for local dosimetry and radiation sensitive field effect transistor (RadFETs) (metal oxide semiconductor (MOS) structures sensitive to ionizing radiation) as real-time dosimeters (Bühler *et al.*, 1996).

10.18.4.18 Scanning Tunneling and Atomic-Force Microscope

A scanning tunneling microscope (STM) is used to image surfaces of electrically conducting materials at atomic scales. An atomically sharp metal tip is brought very close to the surface and a voltage is applied between the sensing tip and the sample, driving a small tunneling current whose magnitude depends on the distance between the tip and the surface. As the tip is moved laterally across the surface, a feedback mechanism moves the tip up and down to maintain a constant tunneling current. 2-D raster scanning with the tip across the surface produces a topographic map of the surface.

An atomic-force microscope (AFM) works on a similar principle, but is not restricted to electrically conductive materials. The AFM tip is sensitive to atomic and intermolecular forces (such as van der Waal forces) and moves up or down on its supporting cantilever. The movement of the tip is monitored by optical means with a laser beam reflected from the tip and read out with a photodiode array or by piezoresistive elements, which follows the movement of the tip and is then read out by a Wheatstone bridge with picometer resolution. Raster scanning the tip across the surface produces a topographic map of the surface with atomic resolution. Piezoceramics are used for the 3-D positioning of the tip at high precision. A major difficulty of STM and AFM for space application is the sample preparation and sample positioning in front of the tip, which requires complicated and sensitive mechanisms. The Midas instrument is an AFM flying on Rosetta (Riedler *et al.*, 2006).

10.18.4.19 Scanning Electron Microscope

A scanning electron microscope (SEM) provides high-resolution 3-D-like images of a sample surface. An electron beam (few hundred electronvolts to 50 keV) thermionically emitted from a tungsten or lanthanum hexaboride cathode or alternatively by field emission (FE) is accelerated toward an anode and focused by a magnetic lens into a beam with a focal spot the size of a few nanometers. Electromagnetic coils are used to steer the electron beam in a scanning mode over a rectangular area of the sample surface.

The electrons hit the surface and are inelastically scattered by atoms in the sample. The scattering process spreads the focused beam up to a few microns onto the surface of the sample leading to emission of electrons and X-rays. The low-energy (<50 eV) secondary electrons can be detected by a scintillator-photomultiplier device and, due to the raster scan a 2-D intensity distribution can be displayed as an image. The number of secondary electrons reaching the detector is translated into the brightness of the image. Emitted X-rays can be analyzed with dispersive X-ray spectroscopy. Also the backscattered electrons can be detected and analyzed as an electron backscattered diffraction image to determine the crystallographic structure of the sample surface.

The spatial resolution of an SEM depends on the size of the focused electron spot defined by the electromagnetic coil system and the extension of the volume in which the electron beam interacts with the sample material. The resolution of the SEM is not sufficiently high to image down to the atomic scale. A miniaturized (11.9 kg) design of a spaceborne SEM has been developed for the (cancelled) Comet Rendezvous Asteroid Flyby (CRAF) Mission ([Albee and Bradley, 1987](#)).

10.18.4.20 Plasma and Wave Instruments

Plasma, the most common phase of matter in the universe, consists of partially or fully ionized gas. It contains an equal number of positive and negative charge carriers, hence behaves quasi-neutral in the stationary state. There are quite a number of different geophysical plasmas with a wide spread in their characteristics, ranging from lower density and lower temperature to high-density and high-temperature plasmas ([Baumjohann and Treumann, 1996](#)). Basically there are two set of instruments required for plasma studies either investigating the

associated electromagnetic waves (plasma wave analyzers (PWAs)) or the charge carriers of the plasma (plasma analyzer).

10.18.4.20.1 Plasma wave analyzer

PWAs measure the electrostatic and electromagnetic components of plasma waves. The instrument receiver is sensitive to the wavelengths from several hertz to several hundred kilohertz in some case to the megahertz range. PWAs are used for a wide range of plasma investigations, for example, magnetospheres, ionospheres, solar wind, atmospheric lightning detection, or events when dust particles strike the spacecraft.

Plasma particles are bound to the natural magnetic field. Motions within the plasma will cause changes with time of the electric and magnetic fields detectable as plasma waves. An electric dipole antenna is used to measure the electric-field component in the plasma, while (triaxial) search coils are used for the magnetic-field components. In the case where very low electric-field frequencies have to be analyzed long dipole antennas are required, sometimes exceeding 50 m single tip length. Such booms are usually designed as wire booms and accommodated on spinning spacecrafts providing gyroscopic stiffness to the wire booms. The deployment of such booms can be problematic and risky for the spacecraft.

The antenna signal is amplified with wideband electrometer pre-amplifiers or a set of band-limited pre-amplifiers for frequency-selective analysis. Low noise design and high impedance with guard feedback systems are required for low density and weak field analysis to minimize the disturbances of the environmental plasma and to achieve enough sensitivity. The pre-amplified signals are usually further amplified for frequency analysis with either discrete filter banks or computational Fourier analysis using digital signal processing (DSP). The same DSP and analog-to-digital conversion chain can be also used for the magnetic field search coil system. Nearly simultaneous measurements of the electric and magnetic field spectrum allow electrostatic waves to be distinguished from electromagnetic waves.

Examples of PWAs are the Wide Band Plasma Wave Investigation for the Earth's Magnetosphere on Cluster II ([Gurnett et al., 1997](#)), the Low Frequency Plasma Wave Analyzer (LFA) onboard the PLANET-B spacecraft for Mars ([Matsumoto et al., 1998](#)) or the Huygens Permittivity, Waves and

Altimetry Analyzer (PWA) for the atmosphere of Titan.

10.18.4.20.2 Plasma analyzer

Plasma analyzers are used to assess the elemental and isotopic composition, the associated energy levels (temperature), density, velocity, and 3-D distribution of interplanetary, planetary ionospheric and magnetospheric, and of the solar wind plasma. They are further used to determine the sources of the magnetospheric plasma, investigate the interaction of ambient plasma with moons and planets, study magnetosphere dynamics, co-rotational forces, field-aligned currents, and magnetic reconnection and the nature of the equatorial current sheet of magnetospheres.

Plasma analyzers are in most cases combinations of instruments to satisfy the wide scope of measurements:

An electrostatic analyzer (ESA) is used to measure the deflection of the ions (and electrons) in a static electric field to define their E/q ratio (energy (E) over charge (q)).

A TOF section is used to measure TOF of the (static pre-accelerated) ions and their energy with a SSD (see Solar Wind Ion Composition Spectrometer (SWICS) on Ulysses; [Gloeckler et al. \(1992\)](#)). From the ESA and TOF measurements simple equations can be used to estimate the energy E , charge state q , and mass m of the ions.

An electron analyzer system (EAS) is used to define the energy of electrons (typically up to a few kiloelectron volts). It consists of two or three combined sensor heads and an ESA with electrostatic deflection (e.g., SWOOPS on Ulysses; [Bame et al. \(1992\)](#)). See also ASPERA-3 on Mars Express, which measures ions, electrons, and high-energy neutral atoms in the outer atmosphere of Mars.

10.18.4.20.3 Langmuir probe

A Langmuir probe is used to determine the electron temperature, electron density, and plasma potential, by inserting one or more electrodes (segmented Langmuir probe) into the plasma, with a time-varying electric potential between the electrode and the surrounding plasma. The measured currents over the applied potentials (I-U curve) allow the determination of the main physical properties of the plasma. Although the measurement itself is rather simple, the interpretation of the data requires understanding of the Langmuir probe theory for a

spherical probe in an unmagnetized plasma (see [Schott, 1968](#)).

The instrument consists of a short boom (to overcome the Debye length and associated plasma distortions) with the sensitive electrode and guard system. The instrument controller applies a time-varying potential to the electrode and measures with a current-sensitive amplifier the associated current to or from the plasma (the current direction depends on the relative applied potential).

One example is the ISL Langmuir Probe Experiment onboard DEMETER ([Lebreton et al., 2006](#)).

10.18.4.20.4 Electron drift instrument

An electron drift instrument (EDI) measures the displacement of a weak ($<1\ \mu\text{A}$) artificially injected electron beam in an ambient electromagnetic field at different electron energies and beam directions. The electromagnetic field introduces a drift of the electron beam. A detector on the opposite side of the spacecraft detects the beam only at some specific geometric and energetic conditions, dependent on the ambient electromagnetic field strength, the angle between injection and magnetic field vectors, and injection energies of the electron beam. By employing two beams and two detectors, these directions can be monitored continuously and the displacement obtained by triangulation. For small magnetic fields the triangulation degenerates and the displacement is obtained instead from the difference in the traveltimes of the electrons in the two beams. Data from magnetometer measurements (fluxgate and search coil) combined with the EDI time of flight analysis and triangulation is used to constrain the ambient electric field. A key advantage of the EDI technique is that the beam probes the ambient electric field at several kilometers away from the spacecraft and hence under undisturbed conditions. As a byproduct, the measured TOFs provide a precise measurement of the magnetic field magnitude ([Paschmann et al., 1997](#)).

The narrow ejected electron beam must be controlled within a hemisphere. Electrostatic multipoles are used to deflect the electron beam in the desired direction. Also the detector has a selectable direction over a wide range. A complex optic in combination with an annular MCP is used to adjust the observational direction of the detector ([Paschmann et al., 1997](#)). A scanning algorithm employing various emission and detector directions combined with

various electron beam energies and beam currents are used to find conditions where the beam returns back to the detector via the ambient field after single or multiple gyrations. The beam, produced by a tungsten filament, is pseudo-noise coded to allow for TOF analysis by correlation (Nakamura *et al.*, 1989).

10.18.4.20.5 Active spacecraft potential control

Measurements of electric field potential may be subject to distortions from interaction of the spacecraft with the plasma and by photoelectrons. Typical spacecraft floating potentials make it in many cases practically impossible to measure the cold ambient plasma (at several electron volts). To mitigate such interactions an active spacecraft control might be required. A Langmuir probe (or similar instrument) is used to measure the spacecraft potential and an ion source is used to compensate for charging by emission of an ion beam. A liquid-metal-ion source has been used for Active Spacecraft Potential Control (ASPOC) on Cluster II (Riedler *et al.*, 1997). A solid needle usually made from tungsten is mounted in a reservoir with heated melted indium (about 520 K). A potential of several kilovolts is applied between the needle and an extractor electrode, which causes the liquid metal to be pulled toward the extractor electrode. Positively charged ions are evaporated in the strong apex field of the needle and the hydrodynamic flow causes liquid metal from the reservoir to follow. The controlled emission of the ions is finally used to reduce the charge of the spacecraft.

10.18.4.21 Magnetometer

Magnetometers are used to map magnetospheres and analyze their dynamics and measure fluctuations in the ambient magnetic field. Magnetometers are usually mounted on the tip of long booms to reduce the magnetic background introduced by the spacecraft and its components. However, not all these effects can be eliminated by separating the instrument by a boom. An additional magnetometer (inbound magnetometer) mounted at a specific boom length fraction can help to distinguish spacecraft-produced fields from ambient fields. In case of high-resolution measurements the spacecraft has to undergo a magnetic cleanliness program, with specific design rules and material selection constraints which introduces cost issues. Another source of potential error in measurement comes from bending

and twisting of the long magnetometer boom. Increase of the boom rigidity (mass increase) or measurement of the actual bending (e.g., with a small stellar compass) can be used to mitigate the influence. In some cases the use of an active boom bending control (smart structures) can be applied. The Ulysses magnetometer uses two sensors representing the two major measurement principles: a fluxgate and a vector helium magnetometer (Balogh *et al.*, 1992). The magnetic fields of the planets and techniques of interpreting magnetic field measurements are discussed in Chapter 10.07.

10.18.4.21.1 Fluxgate magnetometer

A fluxgate magnetometer (FGM) sensor consists of three coils wound around a ferromagnetic ring core, with the primary and secondary windings aligned within the plane of the sensor core, and an additional calibration winding. The primary windings are driven by a current pulse train at a frequency f_0 such that the core is in its region of saturation. In the presence of an external magnetic field, a second harmonic (at the frequency $2f_0$) is generated in the pickup coil (secondary winding), with its amplitude proportional to the co-aligned component of the external field. The phase relation between the pickup signals and the primary drive signal represents the direction of the magnetic field component.

Usually, three identical single-axis fluxgate sensors in an orthogonal arrangement are used to measure the three orthogonal components of the magnetic field. FGMs provide in general sufficient sensitivity, but are subject to some long-time drift, which has to be compensated.

10.18.4.21.2 Vector helium magnetometer

Vector helium magnetometers are based on the influence an ambient magnetic field has on the efficiency with which a metastable population of Helium gas in the triplet ground state can be optically pumped (Zeeman effect). Metastable helium is generated in a helium-absorption cell by RF discharge. An incident radiation produces a distribution of atoms in nonthermodynamic equilibrium among the various energetic substates by optical pumping. Circularly polarized light of wavelength $1.08 \mu\text{m}$ is used for the pumping. The presence of an ambient magnetic field affects the pumping efficiency, which is exploited in the vector mode of the magnetometer: an infrared detector senses the changes in absorption in the helium cell. For a constant light output the change in pumping efficiency is measured, which

depends on the magnetic field strength and angle between the magnetic field vector with respect to the optical axis of the system.

In the scalar mode IR radiation and an AC magnetic field are simultaneously applied to the absorption cell. The efficiency of the optical pumping is minimal and the absorption is greatest when the frequency of the applied AC field is at the Larmor frequency (precession frequency of the magnetic moments of electrons, atomic nuclei, and atoms around the direction of an external magnetic field), which is directly proportional to the ambient magnetic field. Vector helium magnetometers are very sensitive and tend to less drifting compared to flux-gate magnetometer.

10.18.4.21.3 Search coil

Search coil magnetometers are based on Faraday's law of induction and used to measure the AC magnetic field. Several turns of conducting wire are wound around a highly permeable ferromagnetic material core. A changing magnetic field passing through the coil generates a voltage, which is proportional to the time variation of the magnetic field. The sensitivity of search coil magnetometer can be influenced by the area of the coil, the number of turns, the permeability of the core and the rate change of magnetic flux. Specially designed low-noise and low-power amplifiers are used to read out the induced voltage from several hertz to several kilohertz ([Coillot et al., 2006](#)).

10.18.4.22 Accelerometer and Seismometer

Accelerometers are used to sense acceleration of vehicles such as, rovers, aerobots, descent probes, landers, and on board spacecrafts to:

- derive atmospheric parameters (density, drag) during descent/entry phases,
- spin-rate detection,
- derive nongravitational perturbations acting on the spacecraft to determine the actual gravitational field in combination with radio science experiments,
- investigate seismic events, and
- perform inertial navigation.

During the entry and descent phase of an entry probe the interaction with the atmospheres causes drag on the probe, leading to a strong deceleration, which can be measured with accelerometers. Out of the

measured data an estimation of the atmospheric density can be derived.

A spin is often introduced for the probe stabilization during the descent. The spin rate can be measured with accelerometers. This is, beside the engineering aspects important as auxiliary information for scientific measurements performed during the descent ([Lorenz, 2004](#)).

The direct solar radiation and the thermal emission from the planetary surfaces exert a force upon an orbiting spacecraft that, if not accounted for, destroys the accuracy of orbital parameter determination (cf. spacecraft tracking). Very sensitive three-axis accelerometer are used onboard spacecrafsts to separate these nongravitational components ([Iafolla and Nozzoli, 2001](#)).

Various types of accelerometer exist based on the following main principles:

- piezoelectric effect,
- capacitive effect,
- Micro-Electro-Mechanical Systems (MEMS),
- electromechanical servo principle.

Piezoelectric accelerometers use a proof mass, which causes a shear stress to the sensing piezoelectric crystals due to the acceleration. This stress results in a proportional electrical output of the piezoelectric material. With this technique rather simple accelerometer can be constructed. A drawback is the thermal sensitivity of the crystal material.

Capacitive accelerometers use a spring-suspended proof mass, mounted between two capacitor plates. Any acceleration of the proof mass causes a small displacement, leading to a capacity change of the arrangement. The capacitive sensor is applied as frequency defining component in an oscillating circuitry (with a typical bias frequency in the order of 10 kHz; used to reduce the low-frequency Flicker noise contribution of the electronics). Any acceleration of the capacitive sensor causes a frequency modulation. The oscillator output signal is read out via a bridge arrangement with a low-noise amplifier and demodulated to the base band. The minimum detectable acceleration of such an accelerometer is limited by the Brownian noise of the proof masses (and hence its temperature) and by the amplifier noise ([Iafolla and Nozzoli, 2001](#)). Thermal control of the sensor unit within at least 1°C is required for such a system to allow for extremely low noise performance.

10.18.4.22.1 Seismometer

Seismometers are specialized very-low-noise accelerometers to detect seismic waves in planetary bodies. Seismic waves can be used to map the interior of planets, measure and locate quakes and other motions of the lithosphere (*see also* Chapters 10.03 and 10.02, the latter for internal structure modeling).

To define the propagation direction of a wave usually a three-axis seismometer is required. A servo principle is used: a proof mass, confined by electrical forces, is kept at the equilibrium position by the control electronics in a closed loop arrangement. Any vibration of the proof mass, detected by capacitive, inductive, or optical means, is counteracted by the control electronics. Any change of the required electrical force is recorded as the output signal of the seismometer.

For sophisticated seismic analysis a seismometer network is required to constrain the origin, location, and propagation paths of the seismic waves, to disentangle surface waves from interior waves, to distinguish compression (P-waves) from shear waves (S-waves), and finally to constrain the interior structure of a planet or planetary body and its dynamic processes. Such a seismic network requires deployment from multiple planetary landers (cf. NetLander).

So far only the Apollo Seismic Network has returned useful seismic information from the Moon (Lognonne, 2005). Miniaturized seismometers for planetary application are under development (e.g., Banerdt and Pike, 2001). Seismology as a tool for planetary exploration is discussed in Chapter 10.03.

10.18.4.23 Thermometer

Thermometers are used to measure the atmospheric temperature during entry and decent phase and at stationary landers and moving rovers. They can also be used to define *in situ* temperatures of planetary surfaces and are used to derive subsurface temperature, for example, when integrated into a mole or drill system. Thermometers also play an important role as housekeeping sensors to monitor the temperature of instruments platforms or other spacecraft components.

Various types of thermometer principles are in use:

- electrical resistance thermometers;
- thermistors;
- thermocouples; and
- microbolometers.

The basic principle of resistance thermometer is that the conductivity of most conductors and semiconductors strongly depends on their temperature (in most cases a nonlinear dependence). By applying a small constant current the resulting voltage over the resistor can be measured and with proper calibration the temperature derived. Such a sensor principle has been used, for example, to derive the precise atmospheric temperature during the decent of the Huygens Probe in the atmosphere of Titan (Fulchignoni *et al.*, 2005).

Thermocouples are used to measure temperature differences based on the thermoelectric or Seebeck effect. They can be used for wide temperature measurements, but are rather limited in accuracy.

A microbolometer is an uncooled thermal sensor with very high resolution and sensitivity to IR radiation, hence a remote-sensing thermal sensor. The microbolometer reacts with a resistance change when exposed to IR radiation. This resistance change is measured and processed into temperatures. Microbolometer can be made very small and arranged in an array configuration allowing for thermal imaging – even without cooling of the detector array.

10.18.4.24 Pressure

Pressure sensors are used in atmospheres and exospheres to measure *in situ* the static pressure profile, derive the density profile during atmospheric descent, and to measure the diurnal and seasonal changes of the ambient pressure on static or mobile landers and rovers. Two major sensor principles are in use:

- semiconductor strain gage and
- capacitive systems.

The semiconductor strain gages are thin Si-based membranes which change their resistance due to deformation of the membrane caused by pressure changes. The piezoresistive changes of the detector are sensed in a bridge configuration with reference resistors to eliminate thermal drift. The very small bridge output voltage (e.g., 100 mV/350 mbar) is proportional to the sensed pressure and has to be amplified with a low noise differential instrumentation amplifier.

Capacitive systems are based on the distance change of thin capacitor membranes (similar principles as a condenser microphone), which result in a capacitance change. This change is usually sensed in

an oscillator circuit where the sensor capacitor is used as a frequency-defining component. Hence any change in pressure results in a change of the output frequency, which can be easily and precisely measured with a digital counter and precision timing gate.

The pressure sensors are used in three different measurement configurations:

- Absolute pressure,
- Differential pressure, and
- Gauge pressure.

For absolute pressure measurement the sensor membrane is working against a reference pressure (e.g., vacuum) in a closed system. When the reference pressure is 1 atm, then the sensor is called a gauge pressure detector. For differential measurements the membrane works in an open system, hence is sensitive to the pressure gradient.

The Pressure Profile Instrument (PPI) onboard the Huygens Probe ([Harri et al., 1998](#)) is an example of an absolute pressure capacitive system. The detector used as a microphone in the Huygens Permittivity and Waves Analyzer (PWA) is a differential pressure system based on a semiconductor strain gauge detector ([Hamelin et al., 2000](#)).

10.18.4.25 Micro-Gas Sensors

Micro-gas sensors are foreseen to be used for *in situ* measurement on aerobots, landers, and rovers to detect and analyze specific constituents and trace gases of planetary atmospheres. Various principles for micro-gas sensors exist. Some are based on miniaturized IR sensors with emitters employing micro-machined Fabry Perot Interferometers, tuned to wavelengths of an absorption line of a certain gas. As the absorption lines are usually very narrow a combination of narrow band emitter with a narrow band detector is generally required. The attenuation of a beam from source to sensor is measured, from which the concentration of the absorbing species is inferred. The sensitivity of absorption spectroscopy depends upon the optical path length of the radiation, the high channel radiance (radiance integrated over the narrow band of interest) and good collimation of the source.

Specific development for the emitter is based, for example, on a microcavity CdHgTe structure (total thickness of which is about 3–5 μm) between two mirrors with IR emitter. The microcavity is pumped with a laser diode (typically 0.6–0.9 μm wavelength),

resulting in a narrow band emission and a beam with of 3–6 μm .

Another development is based on the so-called electronic nose principle, where integrated sensor arrays are employed. Every single sensor is sensitivity tuned to an individual gas component and responds with resistance change when the sensor is exposed to a particular gas. For example, polymer films are used as the detector, which can change their volume when exposed to a particular gas. The volume change causes a resistance change which can be easily measured. Using an array of different modified polymers, creates very specific resistance change patterns, which are used to identify the gas. Electronic neuronal networks can be used to train the sensor to easier recognize a gas or gas mixture.

Micro-gas sensors and electronic noses are miniaturized alternatives to (rather bulky) gas chromatograph mass spectrometer and currently under development for space applications and already widely used in laboratory and commercial applications.

10.18.4.26 Surface Dating

Luminescence dating is used to determine the age of sediments by determining how long ago mineral grains were last exposed to daylight. Typical minerals that are measured are either quartz or feldspar. The amount of structurally unstable trapped electrons in the crystal structure of minerals is used to define the age of the sediment. Trapped charge accumulates in minerals through time, as energy is absorbed from ionizing radiation (e.g., cosmic rays). A prerequisite for age determination based on luminescence dating is therefore to know the natural radiation environment. Through heating or optical stimulation (optical stimulated luminescence (OSL)) the trapped electrons recombine and the energy is released resulting in luminescence. The brightness of the luminescence signal is a measure of the amount of charge trapped, which is a measure of the ionizing radiation received by the crystal since its last exposure to light ([Aitken, 1998](#)). With the knowledge of the ambient radiation dose rate (based either on measurement or modeling) the age can be derived since the age is equal to the ratio of dose over dose rate.

The instrument consists of a heat or light source for stimulation and a photon detector with appropriate filter to detect the luminescence. Access to buried material and sample preparation are difficult to

realize for planetary robotic missions. OLS instruments are widely used in geosciences on Earth although spaceborne applications for use on Mars are now under development (Kalchgruber *et al.*, 2005).

Another approach to surface dating is based on crater counting as a technique for estimating surface ages over a more extended age range (see Chapter 10.06).

On Earth the radiometric dating technique is widely used for surface dating. The same technique has been used also for Moon samples, which have been returned by Apollo and Soviet Luna Missions. Unfortunately this technique is not (yet) feasible for robotic space missions.

10.18.5 Instrument Suites for Exploration

Single instruments as described in the previous section can never alone satisfy all objectives of a planetary exploration mission. In most cases a combination of instruments is needed to provide new results when exploring our solar system planets and their moons. All missions so far have had a more or less complex set of instruments on board. This approach will remain for future exploration missions, simply because most of the instruments need context information and additional data provided by other instruments on board the spacecraft.

In essence all (geo-) scientific investigation performed on Earth would also be of interest on planets, moons, asteroids, and comets, but restriction due to technology and resource limitations (mass, power, volume, data rate, etc.) have limited what has been

achieved up to now and also what can be done in the near future. Due to the high delta-v requirements of planetary missions (see Chapter 10.17) the propellant consumes a large fraction of the overall spacecraft mass, often exceeding 50% or even 60%. Ever shrinking financial space-science budgets limit the size of future explorative missions to small and medium size. Giant missions such as Cassini–Huygens are in the current budgetary environment unlikely for most of the world's space agencies. Selective and carefully considered priorities of the investigations are required. Focused science objectives, use of appropriate spacecraft and payload technology, paired with cautious tradeoffs on mission risk and new technology are needed.

Smaller spacecraft require optimization of the spacecraft subsystems and in particular of interfaces between spacecraft and payload. Overheads on interfaces, implemented to simplify spacecraft engineering, integration and assembly, compete with additional instrumentation and hence with the scientific performance of a mission.

An interdisciplinary view comprising spacecraft engineering, payload engineering, and science needs to be applied. A resource-saving approach is the use of instrument suites tailored to the scientific requirements rather than collections of individual instruments (Figure 5). A careful instrument suite design based on well-defined scientific requirements allows for significant resource reduction and returns high-level scientific results even on smaller exploration missions.

The following section describes examples of instrument combination for various disciplines. The examples provided are intended as illustration and are under no circumstance exhaustive.

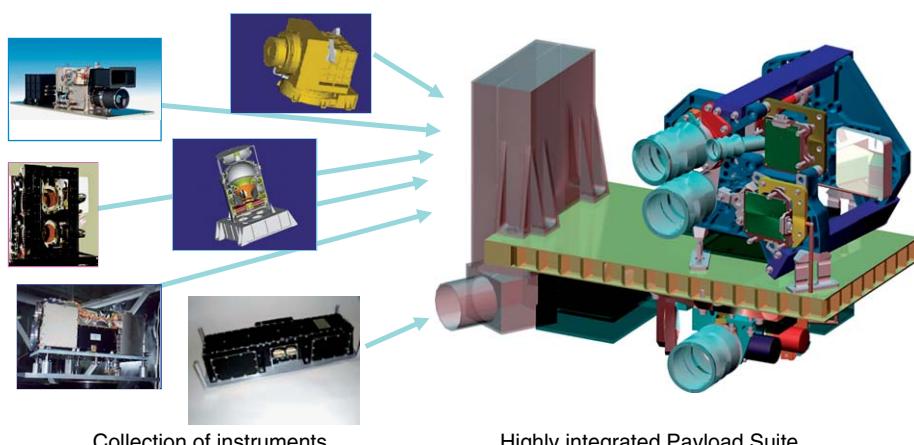


Figure 5 Evolution of collection of instruments into an optimized highly integrated instrument suite (HIPS).

10.18.5.1 Planetary Size, Mass, Shape, and Interior

The precise measurement of a planet's size, its mass, its shape, in particular deviations from the spherical shape by determination of the J-terms (zonal coefficients), localization of mass irregular concentrations (mascons) are fundamental and also required for many other scientific investigations, as well as for precise spacecraft navigation. Interior structure and dynamics, in particular of the terrestrial planets (*see* Chapter 10.02), division of core/mantle/crust and their motion have to be obtained. Although from gravity and magnetic measurement some key information can be derived, seismic data measured on the surface are also important (*see* Chapter 10.03). Of course, landing on a planet is necessary to deploy a seismic station or even preferably a network of stations (*see* Chapter 10.03). Deployment of an instrument suite by means of a high-speed penetrator, as a cost-reducing alternative to soft landing, has been considered (e.g., Deep Space 2 Mars Penetrator or JAXA's Selene – a lunar penetrator), but so far not successfully demonstrated in space. For seismic stations, an extended lifetime of at least several weeks on the planet is mandatory and in many cases very challenging due to environmental conditions and energy limitations of surface stations.

Measurement of the heat flow is important to constrain internal heating and temperatures. Temperatures have to be measured under the surface at various depths with high precision. The development of self-penetrating mole equipped with instrumentation to measure the heat flow is ongoing (Spohn *et al.*, 2001). Magnetometers play an important role by measuring the magnetic field and hence electromagnetic material properties of the interior and constraining the internal field generation. Both measurements from orbit and at the surface are of importance.

10.18.5.2 Surface and Subsurface

Surfaces are analyzed via remote-sensing and *in situ* observation. A long list of interesting phenomena exists, resulting in a huge variety of possible instrumentation. Nevertheless, a kind of standard set of instrumentation has emerged, including imaging and observation in the electromagnetic spectrum (spectroscopy and radiometric measurements), from radio waves to γ -rays, mainly to define morphology and chemical composition. Laser and radar altimeters are used for topographical studies on the surfaces and

Table 2 Instrument examples to constrain planetary size, mass, shape, and interior

Remote sensing	In situ
• Laser or radar altimeter	• Seismometer
• Radio science and accelerometer	• Heat-flow experiment
• Ground penetrating radar	• Magnetometer
• Imager	

to measure gravity tides and surface roughness (**Table 2**). Ground penetrating radar (cf. Mars Advanced Radar for Subsurface and Ionospheric Sounding (MARSIS) on Mars Express; Picardi *et al.* (2004)) is used for subsurface investigations because of the deep-penetration capabilities of low-frequency waves. An interesting application for ground-penetrating radar is the search for subsurface water or to investigate the ice–water interface on Europa, although penetration of the full thickness of the ice layer is rather challenging. High-resolution imaging is important to provide photogeologic details, morphologic and topographic information (particular in combination with stereo imaging and laser or radar topography), and is also used for surface dating based on crater counting.

High-resolution spectral analysis in UV and NIR is used to define the chemical mineralogy possibly enhanced with elemental compositional information provided by XRF and γ -ray spectroscopy. An important factor is a high spatial resolution and the possibility of correlation with images provided by high- and medium-resolution imagers in the visual range, under the restriction that no dense atmosphere prevents optical access to the surface from an orbiting spacecraft. Imaging spectroscopy in the various bands with co-aligned pointing to the visual imager and the star trackers, which are needed to derive the absolute pointing knowledge, can be best realized in an integrated suite approach. The instrument optics and focal planes are mounted very close together on common small platforms (Kraft *et al.*, 2005, Falkner *et al.*, 2004). This avoids structural problems (bending, thermoelastic expansion), which are more difficult to handle when distributed instruments and navigation subsystems are accommodated on an extended instrument platform. A simplified exchange of data from the instrument suite can be easily introduced in the design of the Highly Integrated Payload Suite (HIPS) data handling system.

Magnetometers are used to map magnetospheres and analyze their dynamics and measure fluctuations

in the ambient magnetic field. Most recent developments of a single chip low power ASIC for complete 3-axis fluxgate magnetometer, using higher order sigma-delta conversion and digital filtering (Magnes *et al.*, 2006) demonstrate the possible route for miniaturization of instruments.

Another class of measurements can be done *in situ* on planetary surfaces. Surface access to planets remains challenging due to local environmental conditions (e.g., high temperatures on Venus or Mercury; radiation in the Jovian system), high delta-v requirements (e.g., Mercury, Galilean moons, Titan) and overall technical complexity. Nevertheless many missions have achieved successful landings and new ones are under preparation. Initial stationary landers such as the Viking landers on Mars (Soffen and Snyder, 1976) have been enhanced with mobile rovers (e.g., Mars and Moon), self-penetrating moles, and drills are under preparation (Richter *et al.*, 2005). These approaches require new instrumentation with significant miniaturization and use of very advanced technology, particularly microelectronics. Table 3 lists some of the instrumentation already used, planned, or under development for *in situ* observations.

Table 3 Instrument examples to analyze planetary surfaces and subsurface by means of remote sensing and *in situ* instruments

Remote sensing	In situ
• High-resolution imager (visual)	• Panoramic camera
• Stereo camera	• Laser mass spectrometer
• Spectral imager (UV, NIR, FIR, visual)	• LIBS/Raman spectrometer
• Surface- and subsurface-radar	• Alpha particle X-ray spectrometer
• Laser altimeter	• X-ray diffractometer
• X-ray, gamma ray and neutron spectrometer	• Attenuated total reflection spectrometer
• Microwave sounder, submillimeter sounder	• Mössbauer spectrometer
• Magnetometer	• Infrared Fourier spectrometer
• Thermal infrared radiometer	• Gas chromatograph and mass spectrometer
	• UV spectrometer
	• Microscopic imager
	• Heat flow and physical properties package
	• (Miniature) thermal emission spectrometer
	• Magnetometer
	• Seismometer
	• Ground penetrating radar (surface deployed)

10.18.5.2.1 Surface station

A surface station is a stationary platform deployed on a planetary body, providing the infrastructure (power, thermal environment, data handling, communication, etc.) for the scientific instrumentation (see Table 3 under *in situ*). Deployment mechanisms are required to allow the instruments unobstructed access to the surface and give some degree of freedom for the positioning. Several instruments require very specific alignment with the sample (e.g., LMS, APX, and Mössbauer) and hence a precise deployment mechanism. Imager and panoramic imager should be mounted on a deployable boom to provide an elevated view. A mechanism can provide rotation for elevation and azimuth, allowing for panoramic imaging. A thermometer, wind meter, magnetometer, and many other instruments require some separation from the surface station, simply to avoid the influence and disturbances on their measurements, and are therefore preferable mounted on a (deployable) boom. A robotic arm can introduce a wider range for positioning of instruments (cf. Beagle II; Pullan *et al.* (2004)).

10.18.5.2.2 Rover

An extended range of mobility can be achieved with a rover. A small rover can carry a miniaturized payload suite (e.g., APX and Mössbauer and a microscopic camera). In the simplest form the rover is connected to the surface station via a thin tether for power supply, data transfer, and commanding. Due to the tether the range is limited to several tens of meters (cf. Nanohod; Bertrand *et al.* (2000)).

A free-traveling Rover requires a local power supply (batteries, RTG, solar panel) and a remote link for data transfer and commanding. These subsystems drive the minimum size and mass of the rover (cf. Mars Rover, ExoMars Pasteur), but allowing at the same time for an accommodation of an enhanced payload suite.

10.18.5.2.3 Subsurface access

Access to the subsurface is of interest to understand layering, have access to unaltered material, measure heat flux from the interior of the body, and define soil characteristics. Nonintrusive access is given by means of subsurface penetrating radar or to some limited extent also with a microwave sounder. A drill allows for local subsurface access for either deployment of embedded sensors or to retrieve

samples. Another possibility is the use of a mole (cf. Pluto on Beagle 2). On ice-covered bodies (e.g., comets, Europa, Mars poles) a melting probe can be used for ground penetration by sampling the medium as the probe melts through the ice.

10.18.5.2.4 Mechanical tools

Mechanical tools are used for removal of the top layers of samples to provide access to unaltered pristine material and are often referred to as a ‘geologist’s hammer’. A grinder is a rotating device able to create a hole of a few centimeters in diameter and a few millimeters depth in rocks, which is then further explored with *in situ* instruments (cf. Rock Abrasion Tool (RAT) on Mars Exploration Rover (MER) or the corer-grinder on Beagle 2; Richter *et al.* (2002)). Understanding the differences between the original rock surface and the interior reveals information about the rock formation process and environmental conditions, which have altered the outer layer of the rock.

Another concept is the ultrasonic drill with a piezoelectric driven mini hammer tip, which pulverizes the material for further analysis (Chipera *et al.*, 2003).

10.18.5.3 Atmospheres, Exospheres, and Ionospheres

To understand the complex dynamics of atmospheres and their relationship with the surface an extended set of instrumentation is essential. Also for atmosphere research the combination of remote sensing and *in situ* observations are applied. Balloons, commonly used in the Earth’s atmosphere for *in situ* sounding, are yet rarely used in planetary

atmospheres as instrument platforms (e.g., VEGA 1 and 2 at Venus 1985). Entry and descent probes (e.g., Galileo at Jupiter, Huygens at Titan, and various Venus probes) have delivered significant data due to their capabilities to provide *in situ* compositional and thermodynamic profiles. Lander instruments play a key role in atmospheric measurements because they offer unambiguous observations of diurnal and meteorological variations, in particular for surface-atmosphere interactions, which are difficult or impossible to observe from orbiting spacecraft by means of remote sensing (Morgan *et al.*, 1995). Many open questions remain, due to the dynamic changes associated with atmospheres and many observations over long periods of time are mandatory while comparative interpretations on multiple planets are required. Networks of surface stations capable of long observations, long-duration aerobots cruising in various altitudes, and the deployment of microprobe swarms are important for the next steps in exploration. The key problems with very small stations and microprobes is the maintenance of thermal conditions within hostile environment, the energy supply and communication (in particular energy requirements for communication and data transfer) with the data relay (direct Earth communication is difficult due to energy density limits of current systems). Some key technologies for microprobes are under development such as localization and communication (Wells *et al.*, 2004).

Remote sensing is performed by nadir observations and also limb sounding. **Table 4** lists a few key instruments for remote sensing and for *in situ* investigations of atmospheres. Imagers are important to observe the dynamics of an atmosphere. High resolution is usually not required; a wide field of view is preferred to

Table 4 Instrument examples to investigate planetary atmospheres, exospheres, and ionospheres via remote sensing or *in situ* analysis

Remote sensing	In situ
<ul style="list-style-type: none"> • Wide field imager (visual) • Spectrometer (UV, NIR, FIR, visual) • Laser altimeter • Microwave sounder, submillimeter sounder • Radar including ionospheric sounder • Radio science • Lightning detector 	<ul style="list-style-type: none"> • Aerosol and dust analyzer • Pyrolyzer and mass analyzer • Magnetometer • Nephelometer, miniature gas sensors, and electronic noses • Miniature gas sensors and electronic noses • Anemometer • Up- and down-dwelling flux sensor • Gas chromatograph and mass spectrometer • Pressure, temperature, acceleration sensors • Space plasma and energetic atoms analyzer

guarantee longer-term observations of time-dependent variations and dynamical processes. Spectrometers are used to study the composition in various wavelengths. Atmospheric weighting functions are applied to perform sounding at different altitudes. These functions are describing the layer of the atmosphere, from which the radiation measured at a particular wavelength was emitted (Elachi 1987, pp. 302–311). Measurements performed at various wavelengths reflect peak emissions from different altitudes. Similar principles are used for microwave and submillimeter sounders. Microwave sounders are in particular sensitive to reduced forms of oxygen and nitrogen (water and ammonia). Laser altimeter can be used to define the local atmospheric opacity and atmospheric aerosol loading (Ivanov and Muhleman, 1998).

Radar signal transmission is dependent on the total electron concentration (TEC) in the atmosphere with additional phase shifts and damping occurring in the ionospheres. Both effects are exploited with radar measurement. Precipitation and water vapor content measurements are also classical applications for radar observations. Also limb sounding and observations by occultation (e.g., of telecommunication radio link signals) are exploited by radio science to define, for example, the density profile of an atmosphere.

Lightning detectors are either accommodated on orbiting spacecraft or fly-by spacecraft (e.g., Radio and Plasma Wave Science (RPWS) on the Cassini Orbiter (Fischer et al. (2005)) or by *in situ* instrument such as the HASI Permittivity Waves and Altimetry Analyzer (PWA), installed on the Huygens Probe (Fulchignoni et al., 2005).

A most recent example of a mission dedicated to study the atmosphere and its interaction with the solar wind is Venus Express. Juno is another example of a planned mission to Jupiter containing significant instrumentation for atmospheric sciences.

10.18.5.3.1 Entry and descent probes

Every *in situ* access to a (dense) atmosphere and to the surface of a planet with a (dense) atmosphere requires an entry and descent vehicle. The entry vehicle has to suffer a significant heat flux when entering the atmospheric interface due to the high entry velocities (e.g., around 9.8 km s^{-1} for Venus entry or minimum of 47 km s^{-1} for Jupiter), when performing a hyperbolic insertion. Entry is very challenging in particular for the large gaseous planets with their strong gravity field. A strong, robust thermal protection is required to keep the subsystems and

instruments within an acceptable temperature range during the entry. In some cases the thermal protection system exceeds even more than half of the overall system mass of the entry vehicle. The remaining mass must be shared by the subsystems (structure, power, data handling and communications, etc.), leaving only a small fraction for scientific instrumentation. Miniaturization is a key prerequisite to be able to accommodate a meaningful payload also on smaller systems. Access to the environment is limited during the entry phase as the probe is usually cocooned and thermally fully isolated. Acceleration can however be measured in this cocooned state to define the drag in order to derive the ambient atmospheric density. Thermal sensors and strain gages embedded in the ablative material are employed to understand engineering details like the ablation processes during the entry. After the deployment (jettison) of the aft and front heat shield all scientific instruments can be deployed. As soon as the RF system is operable (in particular after the normal blackout period) important data on the entry trajectory can be obtained exploiting the Doppler shift of the RF-carrier signals. An example for an entry sequence and consecutive deployment of scientific sensors is given by the Huygens Probe, which successfully entered the Titan atmosphere on 14 January 2005 or by the Galileo Entry Probe (Figure 6). Both the Huygens Probe and the Galileo Probe had been designed as descent probes, able to deliver science data during the descent in the atmosphere of Titan and Jupiter, respectively.

The following list summarizes typical instruments used on such kind of descent probes (Table 5).

In many cases the entry probes are optimized to deliver a lander, rover, or aerobot, with a strongly reduced or even no scientific payload at all installed on the probe itself, but only a dedicated payload on the delivered unit.

10.18.5.3.2 Aerobots

Aerobots are aerial robotic platforms (e.g., VEGA balloons) providing an excellent platform for *in situ* sensing instruments at specific altitudes in the atmosphere of planetary bodies. Aerobots have to be delivered by an entry probe before they can be deployed at the target altitude in the atmosphere of interest. Various aerobot principles exist although not many have been flown in space so far. The lighter-than-air balloons (light gas balloons, airships, blimps, Montgolfier) are based on the buoyancy effect using a gas or gas mixture which is lighter

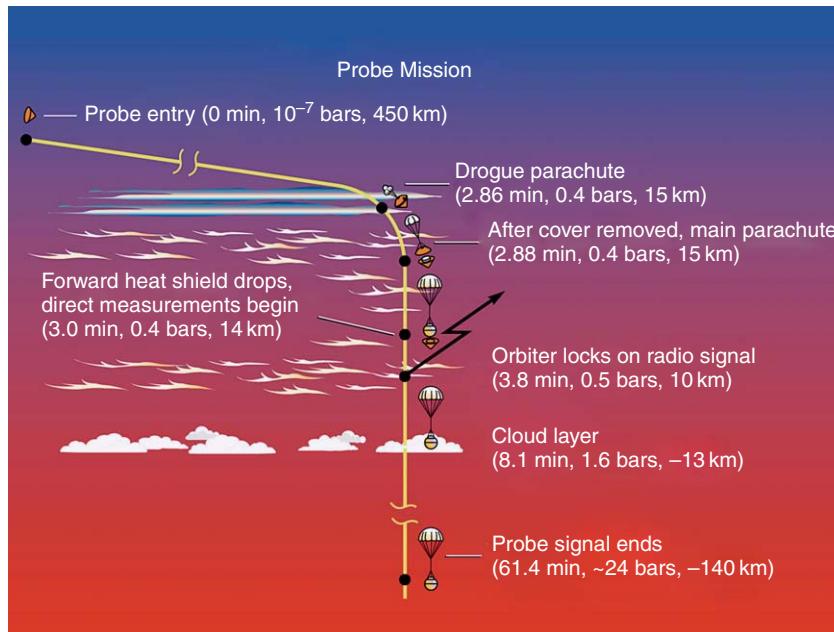


Figure 6 The Galileo Probe Mission events diagram (Image NASA-JPL).

Table 5 Typical payload examples for entry and descent probes in planetary atmospheres

Entry and descent probe payloads

- Atmospheric structure package (accelerometer, gyros, pressure and temperature sensor, recession measurement)
- Miniaturized (spectral) imager
- (Polarization-) nephelometer
- Anemometer
- Up- and down-dwelling flux sensor
- Miniaturized gas chromatograph and mass spectrometer
- Aerosol analysis
- Energetic particles and plasma spectrometer
- Doppler wind experiment (via radio science)
- Acoustic instrument (e.g., to define speed of sound)

composition, horizontal and vertical winds, layering, cells fluxes, density, pressure profiles, gas mixture, elemental sources and sinks, and eventually atmospheric escape processes. Certainly, it is also possible to perform remote observations of the surface, provided the atmosphere is not completely opaque. Payload miniaturization is required as the available payload mass fraction on an aerobot is rather small. The aerobot itself is limited in (undeployed) size and mass to be accommodated within the entry probe, which on its own is already significantly mass constrained due to the high mass fraction of the heat shield and thermal protection system. Aerobots are also a possible platform for deployment of a single or a swarm of microprobes.

than the surrounding atmosphere. Heavier-than-air vehicles (gliders, rotating wings, airplanes) are based on a relative motion through the atmosphere to provide a lift. The main challenges for both principles are the deployment into the atmosphere and thereafter the survival of the local environmental conditions over the required observation time.

Almost all of the *in situ* instruments listed in **Table 4** can also be used on an aerobot in a miniaturized form to measure, for example, atmospheric

10.18.5.4 Magnetosphere

Investigation of the planetary magnetism is about characterization of frozen ambient fields in bodies, which have cooled well below the Curie point and the magnetic fields originating from a planetary dynamo (see Chapter 10.07). Within the planetary magnetospheres, the magnetic field dominates the behavior of electrically charged particles (Kivelson, 1981). Interesting on one side is the interaction of the planetary fields with the solar wind and its magnetic

field, resulting in a bow shock followed by the magnetopause on the upwind side and the long turbulent magnetotail on the downwind side, the magnetically trapped particles in the radiation belts, direct entry of solar-wind particles into the magnetosphere through the polar cusp (aural effects), and interaction of the field with any associated moons. On the other hand the magnetic field provides significant data about the (deep) interior of the planet.

Instrumentation follows these science lines and concentrates on the definition of the internal magnetic field to constrain the interior of the planet or to investigate the magnetosphere and the magnetohydrodynamical (MHD) processes, plasma and waves, reconnection, sub-storms and turbulences in it.

For investigation of the internal field a combination of spacecraft at different altitudes is preferred: one spacecraft on a circular polar orbit to measure precisely the internal field, a second spacecraft on an elliptical polar orbit to understand the interaction with external fields and possibly a surface station, which provides *in situ* reference of the local surface field.

To investigate magnetospheric and plasma processes a tetrahedron formation with four similar spacecraft with identical instrumentation is of use, such as ESA's Cluster-II Mission ([Escoubet et al., 2001](#)). Spinning spacecraft are preferred to accommodate long wire booms, which derive their attitude stabilization by the centrifugal force. The spin also provides 360° scan for all instruments with a radial bore sight without moving mechanism.

Space plasma dynamics involve processes on four scales: (1) the global scale, (2) the MHD scale, (3) the ion scale, and (4) the electron scale. Observation of processes can be best achieved with four spacecraft in a tetrahedron formation like Cluster II, but at least at three scales (MHD, ion, and electron scale) at the same time, thus requiring up to 12 spacecraft in total (cf. MMS Mission, Scope, or Cross Scale). Instrumentation on a particular scale can be identical, but differ between the scales to cope with particular particle types, energy ranges, and dynamics associated with the processes at the different scales ([Table 6](#)).

10.18.5.5 Moons

Payloads for moons are basically similar to what has been described already for their mother planets. Although additional topics such as the interaction of the moon(s) with the mother planet (and their rings), generated tides and associated tidal heating (cf. Io and Europa), gravity perturbations or mutual

Table 6 Examples of Instrumentation for magnetospheres and plasma processes

In situ

- Magnetometer (AC and DC field)
- Plasma particle package
- Electron drift instrument
- Active spacecraft potential control
- Plasma wave analyzer
- Neutral and ionized particle analyzer
- Radiation monitor
- Fast and hot plasma analyzer

influence and interaction with magnetosphere (cf. Ganymede) and particles, (cf. Io; [Spencer et al. \(2002\)](#)) are of particular additional interest. As an exception Titan has, as the only moon in our solar system, a dense atmosphere and hence entry and decent probes with particular tailored payloads, (cf. Huygens; [Lebreton and Matson \(2002\)](#)) are possible ([Table 5](#)).

Due to the lack of a dense atmosphere around the other moons, landing requires full active control and cancellation of the orbital and gravitational velocity with chemical motors. Despite the relative lower gravity of most of the moons, high propellant mass has to be carried on board the spacecraft mainly because of the transfer and insertion delta-v requirements, leaving only a relatively small mass available to the surface payload.

10.18.5.6 Rings

As we know by now, each of the giant planets in our solar system possesses a complex and unique ring system ([Figure 7](#)). Composition, origin and age, characterization of the physical properties, in-ring collision and collision processes with external material (meteorites), evolutionary processes, particle size and distribution are some of the key topics to be addressed in the further understanding of ring systems ([Gordon et al., 2002](#)).

Ring observations usually have to be shared or compete on board the spacecraft with other instrumentation, but can be performed also with Earth-based and space telescope observations ([de Pater et al., 1999](#)). Most of the instruments listed in [Table 3](#) (under remote sensing) can in principle be used for ring observations, of the shepherd satellites and their gravitational interaction with the rings and the central body. Imaging, and in particular close-up imaging, compositional analysis with UV and IR

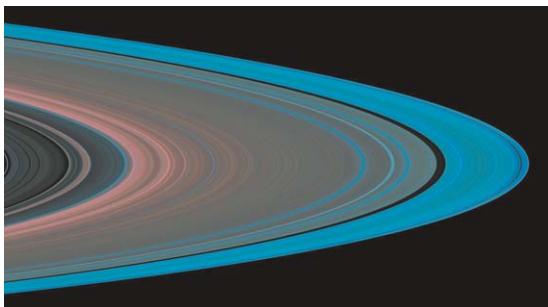


Figure 7 Saturn's ring system observed in the UV by Cassini's Ultraviolet Imaging Spectrograph. The turquoise represents more water ice in the rings (Image NASA/JPL/University of Colorado 'PIA05075: Saturn's A Ring From the Inside Out')

(imaging-) spectrometer play a key role. Long-term observations, required for the understanding of ring dynamics, can be mainly sustained by observation from Earth. *In situ* observations can be done with a dust analyzer, as for the first time performed on Galileo, providing dust spatial densities, grain size distribution, and impact speed along the spacecraft trajectory (Krüger *et al.*, 1999).

10.18.5.7 Asteroids and Comets

Comets represent the most fundamental building blocks of the solar system preserved in a low-temperature and low-gravity environment (e.g., Schulz, 2002). Most of them remain in the Oort Cloud or in the Kuiper Belt outside the orbit of Neptune (*see also* Chapter 10.16). Only occasionally an object is randomly scattered into the inner solar system by gravitational perturbations, becomes active and bright, and can therefore be observed with telescopes from Earth and space or visited by spacecraft (cf. Giotto, Deep Space 1, Stardust, Deep Impact, and Rosetta). The outer comets can hardly be visited with close encounters as the traveltime gets too long and the high escape velocity of the spacecraft necessary for reaching the outer parts of the solar system would require high delta-v maneuvers for trajectory changes to come close to a comet (cf. Chapter 10.17). This is hardly feasible with current propulsion technology. Only those comets crossing the inner solar system on highly elliptical orbits can be visited. While fly-bys are relatively easy to achieve, orbiting missions, such as the Rosetta Mission to comet 67P/Churyumov–Gerasimenko still have high delta-v requirements because the spacecraft has to have a

similar angular velocity as the comet in elliptical orbit around the Sun when getting in orbit around the comet nucleus at a heliocentric distance of about 4.5 AU.

Asteroids have been divided into various classes depending on their distance to the Sun, or more specifically the type of orbit (Chapman, 2005). The asteroids in the main belt are believed to be remnant planetesimals with complex collisional and dynamical activity, whereas NEOs may be either asteroids or dormant comets (Figure 8). They come close to or can even be on Earth-crossing orbits and are therefore potentially hazardous for a collision with the Earth. Tracking and observations from Earth are important for determining the orbits and gross physical characteristics of these bodies, such as size, shape, rotation period, and spectral class, which are important prerequisites for choosing the appropriate targets for exploration missions.

In a fly-by or orbiting mission to small solar system bodies, an accurate determination of their dimensions and rotational properties, as well as their surface, topography, and morphology can be obtained by direct imaging in the visible. The interior can be investigated with ground penetrating radars, like for instance with the Comet Nucleus Sounding Experiment by Radiowave Transmission (CONSERT) experiment (Barbin *et al.*, 1999), a bistatic radar system with the object between transmitter and receiving radar to achieve a tomography of the body.

A radio-science experiment is necessary to define the gravitational field and moments of inertia of the body and for absolute position definition, which is required in relation with altimetry measurement by radar or LIDAR. For the compositional analysis of the surface by remote sensing, multiwavelength spectroscopy is required for wavelengths from γ -rays to submillimeter range. These spectrometers can also be used to determine the outgassing rate of comets and search for activity around asteroids. Surface and subsurface temperatures can be defined with thermal-IR and microwave spectrometers. The long list of possible *in situ* instruments can be used to provide local measurements on the surface and in the case of active bodies, by flying through the coma and tail, determining the composition of the gas and dust particles ejected from the nucleus. By use of drills and moles subsurface characterization can be achieved in a similar style as on planetary surfaces. Table 7 summarizes potential instruments for missions to asteroids and comets.

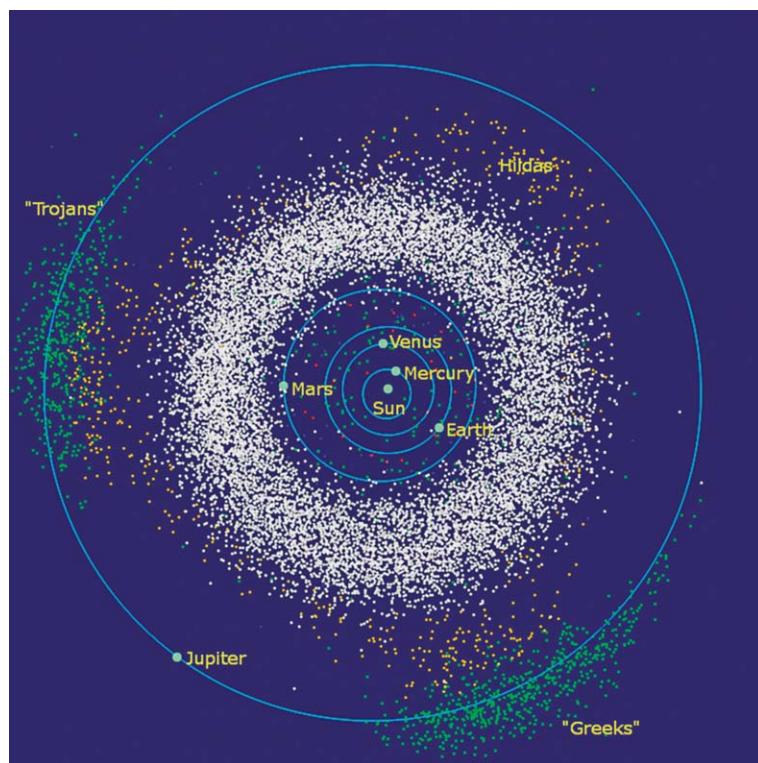


Figure 8 The main asteroid belt concentrates mainly between Mars and Jupiter (Image NASA-JPL).

Table 7 Instrument examples for remote or *in situ* sensing payloads for asteroids and comets

Remote sensing	In situ
<ul style="list-style-type: none"> • Imaging system • Nucleus sounding experiment by radio wave transmission • Mass spectrometers for compositional analysis of neutrals, ions, and dust particles • Radio science experiment • UV, visible and IR imaging spectrometers • γ-ray, neutron and X-ray spectrometers • Microwave sounder • Laser- or radar altimeter • Dust particle analyzer 	<ul style="list-style-type: none"> • Panoramic and microscopic imaging system • Ion mass analyzer • Impact analyzer and dust accumulator • Spectrometer for ion and neutral analysis • Ion composition analyzer • Ion, neutrals and electron spectrometer • Langmuir probe • Magnetometer • Conductivity and permittivity probe • Conductivity probe • Plasma instruments • Alpha X-ray spectrometer • Laser mass spectrometer • Evolved gas analyzer – elemental and molecular composition • Physical surface and subsurface properties • Atomic force microscope • Penetration probes and drills • Seismometer • Acoustic monitor

10.18.5.8 Exobiology

The emerging field of exo- or astrobiology (*see* Chapter 10.12) combines the key characteristics of astronomy, biology, and geology and as such, so does the instrumentation. It comprises most of the instrumentation described so far, but also involves laboratory activities, as well as in-field search of ‘exotic’ life forms on Earth (extremophiles), extending even to the search for extra-solar planets. Within our solar system the focus resides on Mars, Europa, and Titan, due to the presence or assumed presence of water (cf. Europa) and organics (cf. Titan). Search and analysis of organics, search for water as prerequisite for life, search for habitable zones by remote sensing and *in situ* search includes most of the instruments given in **Table 3**, but requires further improvement and new instrumentation, as summarized in **Table 8**. Due to the harsh conditions on most of the planetary surfaces the search is focusing on subsurface niches (e.g., with GPR), requiring mobility and subsurface penetration capabilities (e.g., drills or moles), equipped with sampling mechanism and dedicated sensors for analysis of prebiotic chemistry, as well as for the search for past or present life.

10.18.5.9 Sample Return

A sample return is a very specific case, since the emphasis is on the return of material to be further

analyzed in dedicated terrestrial laboratories, where the same strong resource restrictions are not given as for space mission instrumentation. Not only higher resolution in all measurements can be achieved, but also different techniques applied, which cannot (currently) be used on robotic missions (Sears *et al.*, 2002). The benefits have to be traded with the increased mission complexity, as it requires a spacecraft performing the transfer to the body, soft landing on it (involving either entry where a denser atmosphere exists or active landing on atmosphereless bodies), a dedicate sample mechanism and sample transfer unit, which brings the sample to the container. The container is embedded in the Earth return vehicle (ERV), which does the ascent and escape from the body and return transfer to Earth, where the Earth return container (ERC) has to perform the re-entry and safe landing on the Earth.

In a minimum configuration there will be no particular additional scientific instrumentation foreseen on board the spacecraft, except auxiliary payload for navigation purposes in particular in proximity of the body, where the sample is taken and some minimal instrumentation for sample selection. The focal point will be the sample mechanism and its interaction with the local environment. The easiest case is to take a sample from the atmosphere. More complex is the sampling process for the terrestrial planets, as well as moons, asteroids, and comets,

Table 8 Instrument examples for exobiology payloads suites for remote or *in situ* sensing

Remote sensing	<i>In situ</i>
<ul style="list-style-type: none"> • High-resolution imager (visual) • Stereo camera • Spectral imager (UV, NIR, FIR, visual) • Subsurface radar • X-ray, gamma ray and neutron spectrometer • Microwave sounder, submillimeter sounder • Magnetometer • Thermal infrared radiometer 	<ul style="list-style-type: none"> • Panoramic camera • Laser mass spectrometer • LIBS/Raman spectrometer • Alpha particle X-ray spectrometer • X-ray diffractometer • UV spectrometer • Attenuated total reflection spectrometer • Mössbauer spectrometer • Infrared Fourier spectrometer • Gas-chromatograph and mass spectrometer • Microscopic imager • Heat flow and physical properties package • (Miniature) thermal emission spectrometer • Radon exhalation • Magnetometer • Seismometer • Ground penetrating radar (surface deployed) • Drill system • Life marker chip • Radiation monitor • Permittivity probe • Environmental package

requiring interaction with the surface material (e.g., regolith, ice, rocks).

In a low-gravity environment either an anchoring system is required to provide a stable platform, able to counteract the introduced force by the sample process, or a touch-and-go procedure is applied. Of course a low-gravity environment has the benefit of a simplified return as only a very small delta-v is required to escape from the body. In a high-gravity environment an escape stage is needed to provide enough velocity to leave the body's gravitational field, or at least to return to an orbiter and perform a rendezvous maneuver for sample exchange.

10.18.5.10 Engineering Aspects and Operations

Usually the payload is the main mission driver for cost, mission size, and risk. It is therefore very important to understand and reflect the detailed science requirements and consider all the drivers of these requirements on the spacecraft. The ground resolution of the imaging payload, for example, might be the main driver for the spacecraft attitude control and hence a detailed tradeoff between science requirement and implementation complexity must be performed in an iterative process. A system engineering approach is required to properly manage all interrelations and dependences, requiring interdisciplinary knowledge (*Cruise et al., 1998*).

10.18.5.10.1 Engineering constraints

Resources such as mass, power, data rate, data volume, in particular control on costs strongly limit planetary missions. The selection of instruments must be performed in the first instance so as to satisfy the defined science requirements of a particular mission within the estimated available resources. Use of the latest technology might help to obtain more science per resource, but is not necessarily at lowest risk. A well-considered tradeoff during instrument selection and instrument design is required.

The environmental conditions at the target of observation and also during transfer phases are a strong driver for the design and selection of technology for the complete spacecraft and the instruments in particular. Radiation, thermal ranges, shock and vibration levels are just a few of the drivers, which strongly influence the design solutions of instruments and instrument suites. The payload is commonly the most sensitive part of a spacecraft. Instruments require specific conditions during operation, standby

and at nonoperation (thermal ranges, vibration levels, power supply stability, low-electromagnetic-cleanliness (EMC) noise environment, etc.). Different instruments require different environmental conditions, hence driving the complexity of the spacecraft. In particular the front-end of instruments (optic and focal plain, sensors) have the most stringent requirements.

Co-alignment of the field of view between different instruments and time synchronization of the measurements are additional drivers, requiring stable instrument platforms and precision timing capabilities.

Many low-noise sensors require either active cooling (e.g., peltier element, sorption cooler) or passive cooling via a radiator, because noise increases with temperature. In particular with respect to planets with a hot environment (Mercury, Venus) the thermal design becomes rapidly a strong mission driver.

Sensors requiring direct access to the environment (e.g., thermal sensor, plasma instruments, and magnetometers) must be in many cases mounted outside the protective spacecraft and are consequently exposed to more harsh environmental conditions (radiation, temperature, dust, plumes) or require a boom for deployment. Long appendages such as booms or antennas (in particular for GPR) require mechanism for deployment and change, when deployed, the spacecraft attitude stability, and moment of inertia.

Due to long transfer times (aging), radiation and changing environmental conditions, instruments are subject to performance alteration and drift. A calibration is required and in many cases regular calibration cycles with reference sources have to be performed during the full lifetime of the instrument. Stable calibration targets have to be used for reference and re-pointing of the instrument to the targets requires additional mechanism.

10.18.5.10.2 Operational aspects

All the aforementioned points have to be considered not only in the design of the instrument and instrument suite, but also during operation. The onboard instrument data handling and control computer need to operate the instrument within the required conditions and have to avoid any hazardous conditions. This can be performed either with ground control and commanding or via computer-installed autonomy on board the spacecraft. Data processing has to be performed to produce out of the raw sensor data telemetry packets, which can be transferred via the telemetry link. In many cases the telemetry link is not

available permanently, hence data must be stored onboard the spacecraft to prevent loss of important scientific data. A data management system is required, which takes care of the data acquisition, buffering, and transfer to the telemetry link, when available. Some instruments require already-onboard context information from other instruments to adapt and optimize their own operation. A data exchange and data flow management must be installed in such cases.

The telemetry link is limited in data rate and data volume. Depending on the nature of the data, loss or lossless compression can be used to reduce the data volume of an instrument. For the final data analysis it is important that no compression artifacts are introduced or that these artifacts are properly understood.

Housekeeping data, permanently transferred with science data, summarize the current environmental and status conditions of the instrument, such as temperature, supply currents, voltages, deployment state, pointing elevation, and the current operational mode. This is all of utmost importance to understand later the remotely acquired data set and perform successful data calibration.

10.18.5.10.3 Verification and testing

Not only because of the unique nature of a space mission and the high mission cost, but also because most of the instruments and instrument suites are prototypes, a thorough verification (in particular against the environmental conditions) and testing after the assembly and integration is mandatory. An instrument has to be qualified at harder and verified at least at similar conditions as expected during launch, transfer, and operation of the space mission. Thermal-vacuum testing, vibration testing, radiation testing are just some of the tortures which a space instrument has to survive without damage and serious performance degradation. The particular test procedures and conditions depend finally in the nature of the mission and have to be planned and characterized during the development phase of the mission.

10.18.6 Outlook and Timeline

All our planets (with the exception of Pluto), as well as a number of moons, comets, and asteroids have by now been visited by spacecraft. What remains is a more thorough exploration of all objects to further improve our knowledge and understanding. Landing on planetary surfaces is required to provide ground truth for already existing remotely sensed data (e.g.,

Mercury, Galilean and major Saturn moons, other asteroid classes, and comets). Long-term observations, deployment of meteorological and seismic networks on neighbor planets (Venus and Mars) should also help to better understand our own planet.

Sample and return from the terrestrial planets, various moons, and selected asteroids are further goals to be achieved. Sample and return missions are definitely also technology pre-cursor for long-term future human exploration of, for example, Mars. The further exploration of the interstellar heliopause at some 200 AU from the Sun is another longer-term goal.

Due to the extended duration of space mission development most of the shorter-term and medium-term goals are already under study, planning, or development and embedded in the various national and international space plans. The realization of these plans strongly depends of course on the political environment and the will to financially support these plans.

Technology development is an important prerequisite to prepare further spectacular and striking space missions which are able to add scientific value and allow further exploration. Research and development is definitely a longer-term endeavor, which needs financial support and long-term planning. Technology must be available or in reach when a mission planning starts, to avoid delays, fall-back solutions, cost overruns, or in worst case a mission failure.

Miniaturization and high levels of integration of instruments and spacecraft subsystems similar to what is ongoing in the consumer world, wider use of nanotechnology, systems on a chip, MEMS and further development of space-qualified application-specific integrated circuits (ASIC), new detectors, sensors, materials, and adaptation and qualification of terrestrial instrumentation principles for space application are important midterm steps in preparation of the future of planetary exploration.

Finally, it is our imagination, when coupled to the successful completion and promotion of current space missions, which will prepare the ground for future exploration.

References

- Aitken MJ (1998) *An Introduction to Optical Dating. The Dating of Quaternary Sediments by the Use of Photon-Stimulated Luminescence*. London: Oxford University Press.
- Albee AL and Bradley JG (1987) SEMPA – Scanning electron microscope and particle analyzer for the CRAF mission. *Lunar and Planetary Science XVIII*: 13–14.

- Balogh A, Beek TJ, Forsyth RJ, et al. (1992) The magnetic field investigation on the ULYSSES mission – Instrumentation and preliminary scientific results. *Astronomy and Astrophysics Supplement Series* 92: 221–236.
- Balsiger H, Altweig K, Arijs E, et al. (1998) Rosetta orbiter spectrometer for ion and neutral analysis-rosina. *Advances in Space Research* 21(11): 1527–1535.
- Bame SJ, McComas DJ, Barraclough BL, et al. (1992) The Ulysses solar wind plasma experiment. *Astronomy and Astrophysics Supplement Series* 92(2): 237.
- Banerdt WB and Pike WT (2001) A Miniaturized Seismometer for Surface Measurements in the Outer Solar System. Forum on Innovative Approaches to Outer Planetary Exploration 2001–2020, p. 2.
- Barbin Y, Kofman W, Nielsen E, et al. (1999) The CONSERT instrument for the ROSETTA mission. *Advances in Space Research* 24: 1115–1126.
- Baumjohann W and Treumann RA (1996) *Basic Space Plasma Physics*. London: Imperial College Press.
- Bavdaz M, Peacock A, Beijersbergen MW, and Parmar AN (2002) Development of X-ray optics at ESA. *Proceedings of SPIE* 4496: 94–103.
- Bell JF, Joseph J, Sohl-Dickstein JN, et al. (2006) In-flight calibration and performance of the Mars Exploration Rover Panoramic Camera (Pancam) instruments. *Journal of Geophysical Research* 111: E02S03.
- Benkhoff J, Helbert J, and the MERTIS Team (2006) Thermal infrared spectroscopy to investigate the composition of mercury – The MERTIS instrument on BepiColombo. *Advances in Space Research* 38: 647–658.
- Bertrand R, Brückner J, and van Winnendaal M (2000) The Nanokhod Micro-over – A versatile platform for surface exploration of celestial bodies. In: Foing B and Perry M (eds.), *ESA SP: Exploration and Utilisation of the Moon*, vol. 462, pp. 149–154. ESA SP-462: European Space Agency, ESA Special Publication, Noordwijk, The Netherlands.
- Biccari D, Picardi G, Seu R, Orosei R, and Melacci PT (2002) Mars orbital laser altimeter and mars advanced radar for subsurface and ionosphere sounding (MARSIS). *Proceedings of the SPIE* 4925: 10–17.
- Bohren CF (1983) *Absorption and Scattering of Light by Small Particles (Section 13.7)*. New York: John Wiley and Sons.
- Boynton WV, Feldman WC, Mitrofanov IG, et al. (2004) The Mars Odyssey Gamma-Ray Spectrometer Instrument Suite. *Space Science Reviews* 110: 37–83.
- Brinkhoff WB (2005) On the possible *in situ* elemental analysis of small bodies with laser ablation TOF-MS. *Planetary and Space Science* 53: 817–838.
- Brownlee D, Tsou P, Aleon J, et al. (2006) Comet 81P/Wild 2 under a microscope. *Science* 314: 1711–1716.
- Bühler P, Desorgher L, and Zehnder A (1996) Simple instruments for continuous measurements of trapped particles. *ESA Symposium on Environment Modelling for Space-based Applications*. ESA SP-392, 87.
- Burks M, Cork CP, Eckels D, et al. (2004) Thermal design and performance of the gamma-ray spectrometer for the MESSENGER spacecraft. In: Seibert JA (ed.) *Proceedings of the IEEE Nuclear Science Symposium*, vol 1, pp. 390–394. Piscataway, NJ: Institute of Electrical and Electronics Engineers (IEEE).
- Castagner J-L and Bigioi IJ (2006) Polar nephelometer based on a rotational confocal imaging setup. *Applied Optics IP* 45(10): 2232–2239.
- Chapman CR (2005) Physical properties of small bodies from Atens to TNOs. In: Lazzaro D, Ferraz-Mello S, and Fernandez JA (eds.) *Proceedings of the IAU Symposium 229: Asteroids Comets Meteors*, pp. 3–16. Cambridge: Cambridge University Press.
- Chipera SJ, Bish DL, Vaniman DT, et al. (2003) Use of an Ultrasonic/Sonic Driller/Corer to Obtain Sample Powder for CHEMIN, a Combined XRD/XRF instrument. *Lunar Planet Science XXXIV*: 1603.
- Coillot C, Leroy P, Mosser V, Roux A, and Chanteur G (2006) Wide band search coil magnetometer: A new instrument to investigate magnetic field fluctuations in space. *American Geophysical Union, Fall Meeting*, abstract #SM13A-0334.
- Coradini A, Capaccioni F, Drossart P, Semery A, Arnold G, and Schade U (1999) Virtis: The Imaging Spectrometer of the Rosetta Mission. *Advances in Space Research* 24: 11051114.
- Cruise AM, Bowles JA, Patrick TJ, and Goodall CV (1998) *Principles of Space Instrument Design*. Cambridge: Cambridge University Press.
- De Pater I, Showalter MR, Burns JA, et al. (1999) Keck infrared observations of Jupiter's ring system near Earth's 1997 Ring plane crossing. *Icarus* 138: 214–223.
- Elachi C (1987) *Introduction to the Physics and Techniques of Remote Sensing*. New York: John Wiley and Sons.
- Elachi C, Kuga Y, McDonald KC, et al. (1990) Radar polarimetry for geoscience applications. In: Ulaby FT and Elachi C (eds.), ISBN 978-0-89006-406-1. Norwood, MA: Artech House, Inc.
- Escoubet CP, Fehringer M, and Goldstein M (2001) Introduction The Cluster mission. *Annales Geophysicae* 19: 1197–1200.
- Falkner P, Erd C, and Kraft S (2004) Highly integrated payload suites for planetary exploration. *Geophysical Research Abstract* 6.
- Fischer G, Desch MD, Kaiser ML, et al. (2005) Lightning on Saturn recorded by Cassini/RPWS in the year 2004. *Geophysical Research Abstract* 7.
- Fulchignoni M, Ferri F, Angrilli F, et al. (2005) *In situ* measurements of the physical characteristics of Titan's environment. *Nature* 438: 785–791.
- Funsten HO, Ritzau SM, Harper RW, and Korde R (2004) Fundamental limits to detection of low-energy ions using silicon solid-state detectors. *Applied Physics Letters* 84(18): 3552–3554.
- Garrard TL, Gehrels N, and Stone EC (1992) The Galileo Heavy Element Monitor. *Space Science Reviews* 60: 305–315.
- Giakoumaki A, Osticioli I, and Anglos D (2006) Spectroscopic analysis using a hybrid LIBS-Raman system. *Applied Physics A* 83(4): 537–541.
- Gloeckler G, Geiss J, Balsiger H, et al. (1992) The solar wind ion composition spectrometer. *Astronomy and Astrophysics Supplement Series* 92(2): 267.
- Goessmann F, Rosenbauer H, Roll R, et al. (2006) Cosac, The cometary sampling and composition experiment on Philae. *Space Science Reviews*, doi: 10.1007/s11214-006-9000-6.
- Gommenginger C, Challenor P, Srokosz M, et al. (2006) New scientific applications with Envisat Radar Altimeter individual echoes over ocean, land and ice. *American Geophysical Union, Fall Meeting*, abstract #OS11C-1514.
- Gordon MK, Araki S, Black GJ, et al. (2002) Planetary rings. In: Sykes MV (ed.) *ASP Conference Series: The Future of Solar System Exploration*, 2003–2013, pp. 263–282. San Francisco, CA: The Astronomical Society of the Pacific.
- Gulkis S, Frerking M, Crovisier J, et al. (2006) MIRO: Microwave instrument for Rosetta Orbiter. *Space Science Reviews*, doi: 10.1007/s11214-006-9032-y.
- Gurnett DA, Huff RL, and Kirchner DL (1997) The wide-band plasma wave investigation. *Space Science Reviews* 79: 195–208.
- Hamelin M, Aydogar Ö, Bianchini G, et al. (2000) Surface and sub-surface electrical measurement of Titan with the PWA-HASI experiment on Huygens. *Advances in Space Research* 26(10): 1697–1704.
- Harri A-M, Fagerström B, Lehto A, et al. (1998) Scientific objectives and implementation of the pressure profile

- instrument (PPI/HASI) for the Huygens spacecraft. *Planetary and Space Science* 46: 1383–1392.
- Hartman WK (2005) Adventures in Crater Counting Small Crater Controversies. 36th Annual Lunar and Planetary Science Conference, League City, Texas, abstract no.1427.
- Hauber E and Neukum G (2006) Mars: Simply red? *Astronomy and Geophysics* 47(2): 16–24.
- Huestis et al. (2001) Planetary Atmospheres, Planetary Decadal Study Community White PaperSolar System Exploration Survey, 2003–2012, Executive summary <http://www-mpm.sri.com/decadal/patm-draft.html>
- Huestis DL, Adams NG, Atreya SK, et al. (2002) Planetary Atmospheres. In: Sykes MV (ed.) *ASP Conference Series, Vol. 272: The Future of Solar System Exploration (2003–2013)*, pp. 177–200.
- Iafolla V and Nozzoli S (2001) Italian spring accelerometer (IAS) a high sensitive accelerometer for “BepiColombo” ESA Cornerstone. *Planetary and Space Science* 49: 1609–1617.
- Iess L and Boscagli G (2001) Advanced radio science instrumentation for the mission BepiColombo to mercury. *Planetary and Space Science* 49: 1597–1608.
- Isabelle DB (1994) The pixe analytical technique: Principle and applications. *Radiation Physics and Chemistry* 44: 25–30.
- Ivanov AB and Muhleman DO (1998) Opacity of the Martian atmosphere from the Mars Orbiter Laser Altimeter (MOLA) observations. *Geophysical Research Letters* 25(24): 4417–4420.
- Kalchgruber R, McKeever SWS, Blair MW, et al. (2005) Development of a luminescence dating device for *In Situ* dating of geomorphological features on Mars. *American Geophysical Union*, Fall Meeting, abstract #P41A-0917.
- Kivelson MG (1981) Magnetic Fields and Charged Particles around Major Planets and their Satellites Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences, Vol. 303, No. 1477, The British National Committee on Space Research (Dec. 24, 1981), pp. 247–260.
- Kraft S, Moorhouse J, Collon M, et al. (2005) Demonstration of highly integrated payload architectures and instrumentation for future planetary missions. In: Meynart R, Neeck SP, Shimoda H (eds.) *Proceedings of SPIE 5978: Sensors, Systems, and Next-Generation Satellites IX*, pp. 199–210.
- Kissel J, Glasmachers A, Grün E, et al. (2003) Cometary and Interstellar Dust Analyzer for comet Wild 2. *Journal of Geophysical Research* 108, Issue E10, pp. SRD 4–1.
- Krüger H, Grün E, Heck A, and Lammers S (1999) Analysis of the sensor characteristics of the Galileo dust detector with collimated Jovian dust stream particles. *Planetary and Space Science* 47: 1015–1028.
- Kozyrev AS, Mitrofanov IG, Litvak ML, et al. (2006) The Mercury gamma ray and neutron spectrometer (MGNS) for the ESA BepiColombo mission. *37th Annual Lunar and Planetary Science Conference*, Abstract no.1696.
- Lebreton J-P and Matson DL (2002) The Huygens probe: science, payload and mission overview. *Space Science Reviews* 104: 59–100.
- Lebreton JP, Stverak S, Travnicek P, et al. (2006) The ISL Langmuir probe experiment and its data processing onboard DEMETER: Scientific objectives, description and first results. *Planetary and Space Science* 54: 472–486.
- Lognon P (2005) Planetary seismology. *Annual Review of Earth and Planetary Sciences* 33: 571–604.
- Lorenz RD (2004) Planetary probe entry and descent instrumentation – A review. *Proceedings of the International Workshop on Planetary Entry and Descent Trajectory Reconstruction and Science*, ESA SP-544, 49–56.
- Magnes W, Hauer H, Valavanoglou A, et al. (2006) Magnetometer Front-End ASIC (MFA), 1st International Workshop on Analog and Mixed Signal Integrated Circuits for Space Applications, Xanthi Greece.
- Mahammadzadeh A, Nickson B, and Harboe SR (1998) Dedicated instruments to monitor the space radiation environment. Proceedings of the ESA Workshop on Space Weather, ESA WPP-155.
- Matsumoto H, Okada T, Hashimoto K, et al. (1998) Low frequency plasma wave analyzer (LFA) onboard the PLANET-B spacecraft. *Earth Planets Space* 50: 223–228.
- MEMO WS paper Magnetometer http://www.sciences.univ-nantes.fr/geol/MEMO1stmeeting/Abstracts/24-Cailloit_and_Chanteur-Wide_band_magnetometer-Abs.pdf (accessed Aug 2007).
- Mitrofanov I, Anifimov D, Kozyrev A, et al. (2002) Maps of subsurface hydrogen from high energy neutron detector, Mars Odyssey. *Science* 297: 78–81.
- Morgan T, Abshire J, Clancy T, et al. (1995) Atmospheres from within Planetary Surface Instrument Workshop, LPI Technical Report Number 95-05, Houston, Texas.
- Mottola S, Arnold G, Grothues HG, et al. (2006) The Rosis Experiment on the Rosetta Lander. *Space Science Reviews* (online first) 10.1007/s11214-006-9004-2.
- Nadalini R, Spohn T, Ambrosi R, et al. (2006) The Heat flow and Physical Properties Package (HP3) an integrated sensor suite for planetary subsurface investigations. *Geophysical Research Abstract* 8, 09633, 2006 SRef-ID: 1607-7962/gr/EGU06-A-09633.
- Nakamura M, Hayakawa H, and Tsuruda K (1989) Electric field measurement in the ionosphere using the time-of-flight technique. *Journal of Geophysical Research* 94: 5283–5291.
- Neukum G, Jaumann R, and HRSC Co-Investigator and Experiment Team (2004) HRSC: The high resolution stereo camera of Mars Express. In: *ESA-SP-1240: Mars Express: The Scientific Payload*, 17–35.
- Niemann HB, Atreya SK, Bauer SJ, et al. (2002) The gas chromatograph mass spectrometer for the Huygens probe. *Space Science Reviews* 104: 553–591.
- Paschmann G, Melzner F, Frenzel R, et al. (1997) The electron drift instrument for cluster. *Space Science Reviews* 79: 233–269.
- Paschmann G, Quinn JM, Torbert RB, et al. (2001) The electron drift instrument on cluster: Overview of first results. *Annals of Geophysics* 19: 127388.
- Peg S (1993) Hubble Space Telescope first servicing mission and observatory recommissioning. *Proceedings of SPIE* 1945: 11–16.
- Pellinen R, Piironen J, Silen J, et al. (1990) Laser-ionization studies with the technical models of the LIMA-D/Phobos experiment. *Advances in Space Research* 10: 57–61.
- Picardi G, Biccari D, Seu R, et al. (2004) MARSIS: Mars advanced radar for subsurface and ionosphere sounding. In: Wilson A and Chicarro A (eds.) *ESA SP-1240: Mars Express: The Scientific Payload*, pp. 51–69. Noordwijk, The Netherlands: European Space Agency.
- Pullan D, Sims MR, Wright IP, Pillinger CT, and Trautner R (2004) Beagle 2: The exobiological lander of Mars Express. *ESA SP 1240*: 165–204.
- Richter L, Coste P, Grzesik A, et al. (2005) Instrumented moles for planetary subsurface regolith studies. *Geophysical Research Abstract* 7.
- Richter L, Coste P, Gromov VV, et al. (2002) Development and testing of subsurface sampling devices for the Beagle 2 lander. *Planetary and Space Science* 50(9): 903–913.
- Rieder R, Wänke H, Economou T, and Turkevich A (1997) Determination of the chemical composition of Martian soil and rocks: The alpha proton X ray spectrometer. *Journal of Geophysical Research* 102(E2): 4027–4044.
- Riedler W, Torkar K, Rüdenauer F, et al. (1997) Active spacecraft potential control. *Space Science Reviews* 79: 233–269.

- Riedler W, Torkar K, Jeszenszky H, et al. (2006) MIDAS – The micro-imaging dust analysis system for the Rosetta Mission. *Space Science Reviews*, doi: 10.1007/s11214-006-9040-y.
- Russell EE, Brown FG, Chandos RA, et al. (1992) Galileo Photopolarimeter/Radiometer experiment. *Space Science Reviews* 60: 531–563.
- Sauerbrey G (1959) Verwendung von Schwingquarzen zur Wägung. dünner Schichten und zur Mikrowägung. *Zeitschrift für Physik* 155: 206–222.
- Schott L (1968) Electrical probes. In: Lochte-Holtgraven W (ed.) *Plasma Diagnostics*, pp. 669. New York: Elsevier.
- Schulz R (2002) Comets: Relics of the early solar system. *Journal of Physics IV France* 12/Pr10: 293–305.
- Sears DWG, Allen CC, Britt DT, et al. (2002) Near-Earth asteroid sample return. In: Sykes MV (ed.) *ASP Conf. Ser. 272: The Future of Solar System Exploration 2003–2013* pp. 111–140.
- Singh S (1992) Rowland mounting of holographic ellipsoidal diffraction grating. *Optics Letters* 17(5): 311–313.
- Skolnik MI (1980) *Introduction to Radar Systems*, 2nd edn. New York: McGraw-Hill.
- Soffen GA and Snyder CW (1976) First Viking mission to Mars. *Science* 193: 759–766.
- Spencer JR, Bagenal F, Davies AG, et al. (2002) The future of Io Exploration. In: Sykes MV (ed.) *ASP Conference Series 272: The Future of Solar System Exploration 2003–2013*. pp. 201–216.
- Spohn T, Ball AJ, Seiferlin K, et al. (2001) Flow and physical properties package for the surface of Mercury. *Planetary and Space Science* 49(14/15): 1571–1577.
- Spohn T, Thomas N, Christensen U, et al. (2005) BELA – The BepiColombo Laser Altimeter. *American Geophysical Union, Fall Meeting*, abstract #P41B-0937.
- Willnecker R, Bibring JP, Rosenbauer H, et al., and The Philae Team (2006) Philae (Rosetta Lander): Experiment status after commissioning. *Advances in Space Research* 38: 2025–2030.
- Van Allen JA and Bagenal F (1999) *Planetary Magnetospheres and the Interplanetary Medium, The New Solar System*, 4th edn, 39 pp. Cambridge: Cambridge University Press.
- Wade ML, Agresti DG, Wdowiak TJ, Armendariz LP, and Farmer JD (1999) A Mössbauer investigation of iron-rich terrestrial hydrothermal vent systems: Lessons for Mars exploration. *Journal of Geophysical Research* 104(E4): 8489–8508.
- Wells N, Ballard A, Cosby M, et al. (2004) Atmospheric micropores for Venus: A preliminary probe design and localization method. *Proceedings of the 8th ESA ASTRA Workshop*.
- Wertz JR and Larson WJ (eds.) (1999) *Space Mission Analysis and Design* 3rd edn., Torrance, CA and Dordrecht, The Netherlands: Microcosm Inc. and Kluwer.
- Wright IP, Barber SJ, Morgan GH, et al. (2006) Ptolemy – An Instrument to measure stable isotopic ratios of key volatiles on a cometary nucleus. *Space Science Reviews* (online first).

Relevant Websites

- <http://saturn.jpl.nasa.gov> – Cassini Instrument.
- <http://www2.jpl.nasa.gov> – Basics of Spaceflight.
- <http://www.esa-spaceweather.net> – ESA space weather page.
- <http://www.estec.esa.nl> – ESTEC. 2001
- <http://www2.jpl.nasa.gov> – Galileo Instruments
- <http://newfrontiers.nasa.gov> – JUNO home page.
- <http://lunar.arc.nasa.gov> – Lunar Prospector APS.
- <http://mars.jpl.nasa.gov> – Mars Exploration Rover Mission Instruments Mars Express.
- <http://grs.lpl.arizona.edu> – Mars Odyssey/Mars Odyssey NS.
- <http://mars.jpl.nasa.gov> – Mars Pathfinder.
- <http://www.ivv.fraunhofer.de> – Mass spectroscopy.
- <http://messenger.jhuapl.edu> – Messenger.
- <http://nssdc.gsfc.nasa.gov> – NASA Books on Planetary and Lunar Science.
- <http://sse.jpl.nasa.gov> – NASA Solar System Exploration.
- <http://nssdc.gsfc.nasa.gov> – Planetary Fact Sheets.
- <http://www2.jpl.nasa.gov> – The Magellan Venus Explorer's Guide.