

LECTURE NOTES IN PHYSICS

Sunday (16)

Dynamics of Extended Celestial Bodies and Rings

Taking both a theoretical and observational perspective, this book is an introduction to recent developments in the field of celestial mechanics. It emphasizes the application to extended celestial bodies and devotes much attention to rotational aspects. In particular, it explains the state of art for accurate modelling of the rotation of celestial bodies such as the Earth, the Moon, and Mercury, which involves principles related to hydro-mechanics and geodesy. Comparisons between the light curves of the asteroids and their rotational state are made, and a detailed

Lecture Notes in Physics

Editorial Board

R. Beig, Wien, Austria
W. Beiglböck, Heidelberg, Germany
W. Domcke, Garching, Germany
B.-G. Englert, Singapore
U. Frisch, Nice, France
P. Hänggi, Augsburg, Germany
G. Hasinger, Garching, Germany
K. Hepp, Zürich, Switzerland
W. Hillebrandt, Garching, Germany
D. Imboden, Zürich, Switzerland
R. L. Jaffe, Cambridge, MA, USA
R. Lipowsky, Golm, Germany
H. v. Löhneysen, Karlsruhe, Germany
I. Ojima, Kyoto, Japan
D. Sornette, Nice, France, and Los Angeles, CA, USA
S. Theisen, Golm, Germany
W. Weise, Garching, Germany
J. Wess, München, Germany
J. Zittartz, Köln, Germany

The Lecture Notes in Physics

The series Lecture Notes in Physics (LNP), founded in 1969, reports new developments in physics research and teaching – quickly and informally, but with a high quality and the explicit aim to summarize and communicate current knowledge in an accessible way. Books published in this series are conceived as bridging material between advanced graduate textbooks and the forefront of research to serve the following purposes:

- to be a compact and modern up-to-date source of reference on a well-defined topic;
- to serve as an accessible introduction to the field to postgraduate students and nonspecialist researchers from related areas;
- to be a source of advanced teaching material for specialized seminars, courses and schools.

Both monographs and multi-author volumes will be considered for publication. Edited volumes should, however, consist of a very limited number of contributions only. Proceedings will not be considered for LNP.

Volumes published in LNP are disseminated both in print and in electronic formats, the electronic archive is available at springerlink.com. The series content is indexed, abstracted and referenced by many abstracting and information services, bibliographic networks, subscription agencies, library networks, and consortia.

Proposals should be sent to a member of the Editorial Board, or directly to the managing editor at Springer:

Dr. Christian Caron
Springer Heidelberg
Physics Editorial Department I
Tiergartenstrasse 17
69121 Heidelberg/Germany
christian.caron@springer-sbm.com

Jean Souchay (Ed.)

Dynamics of Extended Celestial Bodies and Rings

 Springer

Editor

Jean Souchay
Observatoire de Paris
Av. de L'Observatoire 61
75014 Paris, France
E-mail: jean.souchay@obspm.fr

Jean Souchay, *Dynamics of Extended Celestial Bodies and Rings*,
Lect. Notes Phys. 682 (Springer, Berlin Heidelberg 2006), DOI 10.1007/b106629

Library of Congress Control Number: 2005930445

ISSN 0075-8450

ISBN-10 3-540-28024-3 Springer Berlin Heidelberg New York

ISBN-13 978-3-540-28024-8 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilm or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable for prosecution under the German Copyright Law.

Springer is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2006
Printed in The Netherlands

The use of general descriptive names, registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Typesetting: by the author using a Springer L^AT_EX macro package

Printed on acid-free paper SPIN: 11398165 55/TechBooks 5 4 3 2 1 0

Editor's Preface

About the Dynamics of Extended Bodies and of the Rings

This book is mainly devoted to celestial mechanics. Under the title above we designate the study of celestial bodies that are not considered as point-masses, as they are often in celestial mechanics, in particular, when dealing with orbital motions. On the contrary, we present and analyse in full details the recent theoretical investigations and observational data related to the effects of the extended shapes of celestial bodies.

Some basic explanations concerning the rotation of an extended body are presented as a tutorial. Then, a large position is reserved for the Earth, which obviously is the most studied planet. We find detailed explanations of the internal structure of our planet, for example, the solid crust, the elastic mantle, the liquid outer core, and the solid inner core. The equations governing its rotational and internal motions under various assumptions (presence of layers, hydrostatic equilibrium etc.) are explained, as well as the modelling of its gravity field and its temporal variations.

We also present the recent developments concerning the dynamics of various celestial bodies. Some of them, the Moon and Mercury, are subject to complex rotational motions related to librations, which are explained exhaustively. Other celestial bodies, such as the asteroids, are undergoing permanent investigations concerning the comparisons between observational data, as light curves, and theoretical modeling of their rotation. The dynamics of these small planets considered as non-rigid bodies are explained in detail.

We also make a complete review of the effects of the impacts on planets and asteroids, and more precisely on their rotational and orbital characteristics. The earlier studies concerning this topic the subject of intensive research are presented.

The concluding part of this book is devoted to the dynamics of the rings and a detailed account of the various equations that govern their motions and evolutions.

We hope that this book will serve as a basis for anybody who wants to become accustomed with the dynamics of extended bodies, and also to get the relevant bibliographic background.

The Thematic School of the CNRS at Lanslevillard

This book is the result of a Thematic School organized by the CNRS (Centre National de la Recherche Scientifique) at Lanslevillard (French Alps) in March 2003, in continuation of previous Winter Schools of Astronomy, organized by C. Froeschlé and his colleagues. This school gathered about fifty people interested in the epistemology, as well as the recent developments in the fields of the rotation of celestial bodies (such as planets and asteroids) and of the rings (such as one around Saturn). This school was organized with the financial support of the CNRS by the intermediary of the “formation permanente” (continuing formation).

We are very grateful to Victoria Terziyan, responsible for the Thematic Schools at CNRS, who was deeply involved in the management of the school, as well as to Liliane Garin and Teddy Carlucci (SYRTE, Observatoire de Paris) who were responsible for the organisation.

Observatoire de Paris
November 2005

Jean Souchay

Contents

Spinning Bodies: A Tutorial

<i>Tadashi Tokieda</i>	1
1 Introduction	1
2 Inertia Matrix	3
3 Conservation Laws	6
4 Miscellaneous Examples	7
5 Euler's Equations	11
6 Spinning under No Torque: Euler's Top	12
7 Some Cases of Spinning under Torques: Lagrange's Top	16
8 Kovalevskaya's Top	19
9 Appendix	20
10 Further Reading and Acknowledgement	21
References	21

Physics Inside the Earth: Deformation and Rotation

<i>Hilaire Legros, Marianne Greff, Tadashi Tokieda</i>	23
1 Introduction	23
2 Terrestrial Mechanics and Survey of Some Dynamical Theories	23
2.1 Historical Review	23
2.2 Physical and Mechanical Setup	24
2.3 Classical Theories	26
3 Deformation of a Planet	29
3.1 Historical Review	30
3.2 Elasto-Gravitational Deformation of a Planet	31
3.3 Viscoelastic Deformation of a Planet	39
3.4 Perspectives	46
4 Rotation of a Deformable Stratified Planet	49
4.1 Historical Review	49
4.2 Rotation with a Fluid Core and a Solid Inner Core	50

VIII Contents

4.3 Discussion	60
4.4 Conclusion	62
References	62

Modelling and Characterizing the Earth's Gravity Field: From Basic Principles to Current Purposes

<i>Florent Deleflie, Pierre Exertier</i>	67
1 Introduction	67
2 Basic Principles	68
2.1 Mass and Gravitation	68
2.2 Potential Generated by a Continuous Body	70
2.3 Potential Generated by a Continuous Body in Rotation	71
3 Coefficients Characterizing the Gravity Field	72
3.1 Legendre Polynomials	72
3.2 Spherical Harmonics	73
3.3 Development of the Gravity Field in Spherical Harmonics	73
4 Global Geodynamics	75
5 Orbital Dynamics	77
5.1 Integrate the Equations of Motion	77
5.2 Computing from Space the Coefficients of the Gravity Field ...	79
6 Current Purposes	81
6.1 Combined Gravity Field Models	81
6.2 The New Missions GRACE and GOCE	83
6.3 Towards an Alternative to Spherical Harmonics for Short Spatial Wavelengths	84
7 Conclusion	85
References	86

Asteroids from Observations to Models

<i>D. Hestroffer, P. Tanga</i>	89
1 Introduction	89
2 Lightcurves	89
3 Rotation	90
4 Figures of Equilibrium	95
4.1 Hydrostatic Equilibrium	96
4.2 Elastostatic Equilibrium and Elastic-Plastic Theories	101
4.3 Binary Systems and the Density Profile	103
5 The Determination of Shape and Spin Parameters by Hubble Space Telescope	105
5.1 The FGS Interferometer	105
5.2 From Data to Modeling	107
5.3 Some Significant Examples	109
6 Conclusions	113
References	114

Modelling Collisions Between Asteroids: From Laboratory Experiments to Numerical Simulations

<i>Patrick Michel</i>	117
1 Introduction	118
2 Laboratory Experiments	120
2.1 Degree of Fragmentation.....	121
2.2 Fragment Size Distribution.....	122
2.3 Fragment Velocity Distribution	122
3 Fragmentation Phase: Theoretical Basis	123
3.1 Basic Equations	124
3.2 Fundamental Basis of Dynamical Fracture	125
3.3 Numerically Simulating the Fragmentation Phase	130
3.4 Summary of Limitations Due to Material Uncertainties	131
4 Gravitational Phase: Large-Scale Simulations	131
5 Current Understanding and Latest Results	133
5.1 Disruption of Monolithic Family Parent Bodies	134
5.2 Disruption of Pre-Shattered Parent Bodies	136
6 Conclusions.....	140
References	141

Geometric Conditions for Quasi-Collisions in Öpik's Theory

<i>Giovanni B. Valsecchi</i>	145
1 Introduction	145
2 The Geometry of Planetary Close Encounters	146
3 A Generalized Setup for Öpik's Theory.....	149
3.1 From Heliocentric Elements of the Small Body to Cartesian Geocentric Position and Velocity and Back	149
3.2 The Local MOID	151
3.3 The Coordinates on the b -Plane	153
3.4 The Encounter	154
3.5 Post-Encounter Coordinates in the Post-Encounter b -Plane and the New Local MOID	155
3.6 Post-Encounter Propagation	156
4 Discussion	157
References	158

The Synchronous Rotation of the Moon

<i>Jacques Henrard</i>	159
1 Introduction	159
2 Andoyer's Variables.....	160
3 Perturbation by Another Body	161
4 Cassini's States	163
5 Motion around the Cassini's States	165
References	167

Spin-Orbit Resonant Rotation of Mercury

<i>Sandrine D'Hoedt, Anne Lemaitre</i>	169
1 Introduction	169
2 Reference Frames and Variables Choice	171
3 First Model of Rotation	173
4 Development of the Gravitational Potential	174
5 Spin-Orbit Resonant Angle	175
6 Simplified Hamiltonian and Basic Frequencies	177
7 Conclusion	180
References	180

Dynamics of Planetary Rings

<i>Bruno Sicardy</i>	183
1 Introduction	183
2 Planetary Rings and the Roche Zone	184
3 Flattening of Rings	185
4 Stability of Flat Disks	186
5 Particle Size and Ring Thickness	190
6 Resonances in Planetary Rings	192
7 Waves as Probes of the Rings	198
8 Torque at Resonances	198
9 Concluding Remarks	200
References	200

Index	201
--------------------	-----

List of Contributors

Anne Lemaître

Département de Mathématiques
8, rempart de la Vierge
5000 Namur, Belgium
anne.lemaitre@fundp.ac.be

Bruno Sicardy

Observatoire de Paris
Bâtiment 10
92195 Meudon Cedex, France
Bruno.Sicardy@obspm.fr

Daniel Hestroffer

IMCCE/UMR 8028 du CNRS
77, av. Denfert Rochereau
75014 Paris, France
hestro@imcce.fr

Florent Deleflie

OCA/Gemini
Av. N. Copernic
06130 Grasse, France
Florent.Deleflie@obs-azur.fr

Giovanni Valsecchi

INAF-IASF
via Fosso del Cavaliere 100
00133 Roma, Italia
Giovanni@rm.iasf.cnr.it

Hilaire Legros

Ecole et Observatoire
des Sciences de la Terre
Laboratoire de Mécanique Globale
5, rue René Descartes
67 084 Strasbourg Cedex, France
Hilaire.Legros@eost.ustrasbg.fr

Jacques Henrard

Département de Mathématiques
8, rempart de la Vierge
5000 Namur, Belgium
Jacques.Henrard@fundp.ac.be

Marianne Greff-Lefftz

Institut de Physique
du Globe de Paris
Département de géomagnétisme et
paléomagnétisme
4, place Jussieu
Paris Cedex 05, France
greff@ipgp.jussieu.fr

Paolo Tanga

OCA/Gemini
Av. N. Copernic
06130 Grasse, France
tanga@obs-nice.fr

XII List of Contributors

Patrick Michel

Observatoire de la Côte d’Azur
(OCA)
UMR 6202 Cassiopée/CNRS
BP 4229, 06304 Nice Cedex 4, France
`Patrick.Michel@obs-nice.fr`

Pierre Exertier

OCA/Gemini
Av. N. Copernic
06130 Grasse, France
`Pierre.Exertier@obs-azur.fr`

Sandrine D’Hoedt

Département de Mathématiques
8, rempart de la Vierge
5000 Namur, Belgium
`Sandrine.Dhoedt@fundp.ac.be`

Tadashi Tokieda

Trinity Hall
Cambridge
CB2 1TJ, United Kingdom
`tokieda@dpmmms.cam.ac.uk`

Spinning Bodies: A Tutorial

Tadashi Tokieda

Trinity Hall, Cambridge CB2 1TJ, UK
tokieda@dpmmms.cam.ac.uk

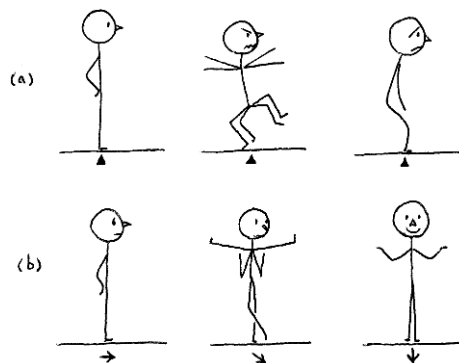
Abstract. The article sets up some of the mathematics underpinning this LNP and is addressed to those who have learnt rigid-body dynamics officially but still feel suspicious toward it. I try to relieve the monotony by discussing unusual examples, and by delving deeper into the usual material than many books.

Contents: 1. Strange rotational phenomena, 2. Inertia matrix, 3. Conservation of angular momentum, 4. Miscellaneous examples, 5. Euler's equations, 6. Euler's top, 7. Lagrange's top, 8. Kovalevskaya's top, 9. Rotational proof of Pythagoras, 10. Further reading.

A Alain Chenciner, maître mécanicien.

1 Introduction

1.1. You are standing on slippery ice. Can you wriggle your body so as to slide and end up standing somewhere else (picture a)?



Well, you can't—no matter how you wriggle yourself, your centre of mass stays on the same spot. Now suppose you try *swiveling* instead of sliding (picture b). Can you end up facing some new orientation? This time you *can*. Stretch out your arms and twist your upper body anticlockwise; your lower body then twists clockwise. Next pull your arms in and untwist your upper body clockwise; your lower body then untwists anticlockwise, *less than it twisted clockwise earlier*. The net effect is, you swivel clockwise by an angle. Denizens of warmer climes may experiment on a swivel chair.

Cats accomplish this feat with instinctive grace: a cat falling upside down twists itself in mid-air and lands upside up, on its paws. I must own that I am too respectful of the feline species to have dared an experiment myself. Instead, here is a design of a cat made of stiff paper. When dropped upside down, this toy cat flips and lands on its paws. (Alas, the physics is unrelated to that of real cats.)

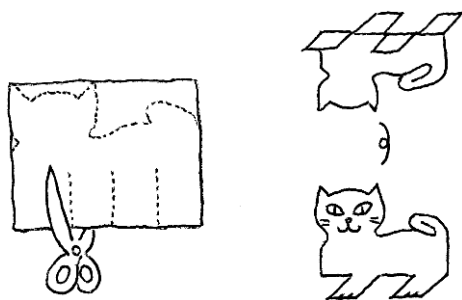


Figure skaters accelerate or decelerate their spin by pulling in or stretching out their limbs.

1.2. Ordinary life offers few opportunities to experience rotational motion. (Never mind for the nonce that we live on a rotating object.) In contrast, translational motion is with us all the time, e.g. when riding a car. But in the days of Galileo & Co., finely controlled translational motion was rare in people's experience; this may explain why dynamics and in particular the law of inertia took long to discover. Controlled rotational motion is not so common to this day, and accordingly dynamics of rotation seems baffling. This article's business is to unbaffle us about dynamics of rotation and to make it as intuitive as dynamics of translation.

1.3. English is rife with pseudo-synonyms of "rotate": "revolve", "spin", "swivel", "turn", "twist", "whirl"... They carry helpful differences of nuance, which we shall turn to our advantage.

1.4. One terminological oddity. Traditionally, rigid bodies are called "tops" (French "toupie", German "Kreisel", Japanese "koma", Latin "turbo", Russian "volchok"). So, from now on,

"Top" and "rigid body" are synonymous,

and a top will be denoted by T . The letter T even looks like a top. Warning: despite the connotation of the word “top”, our tops are not *a priori* assumed to be symmetrically shaped.

2 Inertia Matrix

2.1. *Inertia* is a body’s resistance to acceleration. In translational motion, it is encoded in mass, a scalar: a napping rhinoceros is hard to budge, a charging rhinoceros is equally hard to halt. In rotational motion, resistance is encoded in a quantity more sophisticated than a scalar because, when it is spun about different axes, a body may differently resist rotational acceleration. Rotational inertia turns out to be a *matrix*.

In particle dynamics, mass m appears as coefficient in two quantities: in momentum $p = mv$ and in kinetic energy $E_{\text{par}} = \frac{1}{2}mv^2$ for a particle moving at velocity v . In rigid-body dynamics, the inertia matrix appears also as coefficient in two quantities (2.2, 2.3).

2.2. Given a top T (1.4), imagine rectangular coordinate axes attached to T whose origin is at a point O which may be inside or outside T . The axes as well as O move together with T .

*We always take as O the centre of mass C of the top
or some stationary point (pivot).*

In these coordinates, each point of T is parametrised by a radius vector $x = (x_1, x_2, x_3)$. Let $\rho(x)$ be the density of T at x , $dx = dx_1 dx_2 dx_3$ the volume element.

A top T of mass M is moving at U = velocity of O , Ω = angular velocity around O , so that a point x of T has velocity $U + \Omega \wedge x$ to an observer at rest. The total angular momentum L of T around O is

$$L = \int_T x \wedge (U + \Omega \wedge x) \rho(x) dx = M(C - O) \wedge U + \int_T x \wedge (\Omega \wedge x) \rho(x) dx .$$

The term $M(C - O) \wedge U$ vanishes by our hypothesis that $O = C$ or $U = 0$. The integral term defines an operator, *linear* in Ω hence representable by a matrix, the **inertia matrix** (*alias* inertia tensor) I of T around O :

$$L = I\Omega = \int_T x \wedge (\Omega \wedge x) \rho(x) dx .$$

Thus the first quantity in which the inertia matrix I appears as coefficient: the angular momentum $L = I\Omega$. The dimension of L is mass $\times$ length² $\times$ time⁻¹; that of I is mass $\times$ length².

Note the analogy with $p = mv$ (2.1). Beware however that, because I is a matrix rather than a scalar, in general L is not parallel to Ω . One knack of un baffling ourselves about dynamics of rotation consists in distinguishing clearly between angular momentum L and angular velocity Ω (e.g. 6.2).

2.3. In the scenario of (2.2) and $\langle \cdot, \cdot \rangle$ denoting the scalar product, the total kinetic energy E of T is

$$\begin{aligned} E &= \frac{1}{2} \int_T (U + \Omega \wedge x)^2 \rho(x) dx \\ &= \frac{1}{2} U^2 \int_T \rho(x) dx + \langle U, \Omega \wedge \int_T x \rho(x) dx \rangle + \frac{1}{2} \langle \int_T x \wedge (\Omega \wedge x) \rho(x) dx, \Omega \rangle \\ &= \frac{1}{2} MU^2 + \langle U, \Omega \wedge M(C - O) \rangle + \frac{1}{2} \langle I\Omega, \Omega \rangle . \end{aligned}$$

E splits into two terms, a translational term that has the form as if the mass of T were concentrated at O , plus a rotational term; the cross term $\langle U, \Omega \wedge M(C - O) \rangle$ vanishes by our hypothesis (2.2). Thus the second quantity in which the inertia matrix I appears as coefficient: the rotational kinetic energy $E_{\text{rot}} = \frac{1}{2} \langle I\Omega, \Omega \rangle$.

Note the analogy with $E_{\text{par}} = \frac{1}{2} \langle mv, v \rangle$ (2.1). Beware however that, because I is a matrix, in general E_{rot} depends not only on the magnitude but also on the direction of Ω .

2.4. The inertia matrix I is *symmetric*. Indeed, for any vectors $\Omega, \tilde{\Omega}$,

$$\begin{aligned} \langle I\Omega, \tilde{\Omega} \rangle &= \int_T \langle x \wedge (\Omega \wedge x), \tilde{\Omega} \rangle \rho(x) dx \\ &= \int_T \langle \Omega \wedge x, \tilde{\Omega} \wedge x \rangle \rho(x) dx = \int_T \langle \Omega, x \wedge (\tilde{\Omega} \wedge x) \rangle \rho(x) dx = \langle \Omega, I\tilde{\Omega} \rangle , \end{aligned}$$

which expresses that I equals its own transpose. By a theorem of linear algebra, suitable rectangular axes x_1, x_2, x_3 can be chosen so as to diagonalise I ; they are called **principal axes** (*alias* principal directions) of the top. With respect to principal axes,

$$I = \begin{pmatrix} I_1 & 0 & 0 \\ 0 & I_2 & 0 \\ 0 & 0 & I_3 \end{pmatrix} .$$

The eigenvalues I_1, I_2, I_3 are the **principal moments of inertia**. The **moment of inertia** about an arbitrary axis, without the epithet “principal”, means $\langle Ie, e \rangle$ for a unit vector e along that axis.

One suggestive interpretation of the diagonalisability of I is,

As far as inertial responses are concerned, any top is an ellipsoid.

2.5. Owing to curricula which introduce students to moment of inertia in the context of exercises on multiple integrals, many live under the impression that moment of inertia somehow characterises the mass distribution about an axis. To be sure, it *happens* to be computable from the distribution, but plenty of different distributions result in the same moment of inertia, and anyway mass distribution is not the *raison d'être* of moment of inertia. To repeat, what moment of inertia characterises is the body's resistance to rotational

acceleration. As for mass distribution, the good news about rigid bodies is that details more complicated than the ellipsoid of inertia are invisible to dynamics (cf. form of equations 5.1).

2.6. The shape of a top often makes its principal axes readily identifiable: mentally fit an ellipsoid to the top (2.4). A rectangular box has principal axes parallel to the edges. For a circular cylinder, one principal axis is the axis of the cylinder; the remaining two are *any* two axes perpendicular to the first.

An equilateral triangular lamina is instructive. One principal axis is normal to the lamina. About this axis, the lamina has rotational symmetry of order 3, whereas an ellipsoid with distinct semiaxes admits rotational symmetry of order at most 2. Hence the ellipsoid of the lamina must be *of revolution*, and the remaining principal axes are any two axes perpendicular to the first. In general, as soon as a top has rotational symmetry of order > 2 about some axis, its ellipsoid is of revolution about that axis. If this happens about two axes, then the ellipsoid degenerates to a ball, and any three perpendicular axes are principal.

A quiz. About which axis is the moment of inertia of a cube largest? The axis connecting 1) diametrically opposite vertices, 2) midpoints of diametrically opposite edges, 3) midpoints of opposite faces?

2.7. In desperation I could be computed: unpacking the definition (2.2),

$$I = \int_T \begin{pmatrix} x_2^2 + x_3^2 & -x_1x_2 & -x_1x_3 \\ -x_2x_1 & x_3^2 + x_1^2 & -x_2x_3 \\ -x_3x_1 & -x_3x_2 & x_1^2 + x_2^2 \end{pmatrix} \rho(x_1, x_2, x_3) dx_1 dx_2 dx_3,$$

which reveals again the symmetry of I (2.4). Computing moments of inertia is salutary perhaps for the soul but not for much else; please look them up in your favourite reference. We mention just two tips. First, “Routh’s rule”: the moment of inertia of a homogeneous body about an axis of symmetry is

$$\text{mass} \times \frac{\text{sum of squares of perpendicular semiaxes}}{3, 4, 5},$$

the denominator being 3, 4 or 5 according as the body is rectangular (2D or 3D), elliptical (2D) or ellipsoidal (3D) [14]. Second, if the mass is M and the radius R , the moment of inertia of a homogeneous *solid ball* about its diameter is $\frac{2}{5}MR^2$ (a special instance of Routh), while that of a homogeneous *spherical shell* is $\frac{2}{3}MR^2$ (not an instance of Routh, which does not apply to hollow bodies).

2.8. Faced with a top, our Pavlovian reaction is to think of its moment of inertia around the centre of mass C . Yet it can prove useful to think around other points (e.g. 4.4, 4.5, Sects. 7, 8). The “parallel axes theorem” saves us the trouble of recomputing moments of inertia afresh:

Let I_C [resp. I_O] be the inertia matrix of a top of mass M around C [resp. another point O].

Then $I_O = I_C +$ inertia matrix around O of a particle of mass M at C .

The last matrix may be written $M[{}^t(C - O)(C - O)\delta - \delta(C - O){}^t(C - O)]$, where $C - O$ is a column vector and its transpose ${}^t(C - O)$ a row vector, and δ is the identity matrix (cf. formula of 2.7). The theorem is not used in this article, but it is comforting to know.

3 Conservation Laws

3.1. Dynamics is a drama of conserved quantities: *momentum, angular momentum, energy*. In dynamics of rotation, the star billing goes to angular momentum and rotational energy. All the mathematics we manipulate in this article are auxiliary to them, all the laws we formulate are ultimately about how they do or do not change in time. In every physical problem, we should zoom in on conservation laws: tyros rush to differential equations, whereas pros stick to conservation laws as far as they can.

3.2. A top T of mass M and inertia matrix I around a point O is moving at $V =$ velocity of its centre of mass C and $\Omega =$ angular velocity around O ; our hypothesis (2.2) is that $O = C$ or O is stationary. The momentum and the angular momentum around O of T are $P = MV$, $L = I\Omega$.

*Momentum and angular momentum are conserved,
except for external disturbing influences:*

$$\frac{d}{dt}P = F, \quad \frac{d}{dt}L = N.$$

Here F is the force and N the torque acting on T . If each point x of T is subjected to a field of force $f(x)$, then the total force is

$$F = \int_T f(x)dx$$

while the total **torque** (*alias* moment of force) around O is

$$N = \int_T x \wedge f(x)dx,$$

the radius vector x being measured from O . The dimension of N is mass $\times$ length² $\times$ time⁻², the same as that of energy.

3.3. As everywhere in physics,

Energy is conserved.

Of course our accounting must include all forms of energy: kinetic, potential, heat...

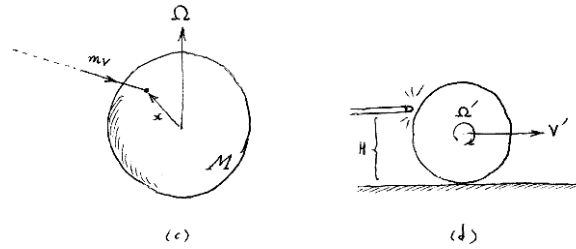
3.4 In many places in the literature, the conservation laws (3.2, 3.3) are “derived” from laws of particle dynamics by regarding a rigid body as an assemblage of particles, etc. Actually it is simpler to adopt the laws (3.2, 3.3) as fundamental in their own right.

3.5. Rigid body is an idealisation, even in macroscopic physics. Relativity teaches that nature knows no such thing as a rigid body. Non-relativistically too, natural matter is more or less *deformable*. Cats actively reposition their bodies and vary their inertia matrices (1.1). A classic example from astronomy is a rotating mass of fluid, e.g. a star; unlike cats, a star passively responds to various forces acting on it and settles into an equilibrium figure. A collection of grains, or rush-hour commuters on the Tokyo underground, can behave like a rigid body or not, depending on how tightly they are packed. This article ignores all these.

3.6. There is almost nothing on rigid bodies in *Principia*.

4 Miscellaneous Examples

4.1. It is remarkable that simple conservation laws (3.2, 3.3) are already amply powerful to solve many nontrivial problems, without further development of formal machinery. In this section we sample several illustrations.



4.2. A meteorite impacts and adheres to a planet. How is the planet's axis of rotation affected (picture c)?

The planet of mass M and moment of inertia I around its centre C is moving at V = velocity of C and Ω = angular velocity around C , when a meteorite of mass m flies in at velocity v and impacts a point x on the planet. Denoting the values after the impact by $'$, we have from conservation of momentum and angular momentum (3.2)

$$\begin{aligned} MV + mv &= MV' + m(V' + \Omega' \wedge x) \\ I\Omega + x \wedge mv &= I\Omega' + x \wedge m(V' + \Omega' \wedge x). \end{aligned}$$

Suppose, reasonably enough, that $m \ll M$, $|v| \gg |V|, |V'|, |\Omega' \wedge x|$. Then the planet's new angular velocity is

$$\Omega' \sim \Omega + \frac{x \wedge mv}{I}.$$

The impact could tilt the axis of rotation appreciably. Perhaps this is the fate that befell Uranus, whose axis of rotation is abnormally tilted from the normal to the ecliptic.

4.3. The next one is a chestnut. If you shoot a billiard ball too high [resp. low], the ball skids with forward [resp. backward] spin. At what height must you shoot so as to induce pure rolling (picture d)?

Assume the motion is restricted to a vertical plane; the problem is then planar. The cue horizontally imparts a force F at height H to a ball of mass M , radius R , moment of inertia $I = \frac{2}{5}MR^2$ around its centre (2.7). Before, the ball had velocity $V = 0$ and angular velocity $\Omega = 0$; after, these will change to V', Ω' , both of which we can leave unknown and yet solve the problem. If the shot occurs during a brief interval Δt , then

$$MV' = F\Delta t, \quad I\Omega' = (H - R)F\Delta t;$$

eliminating $F\Delta t$,

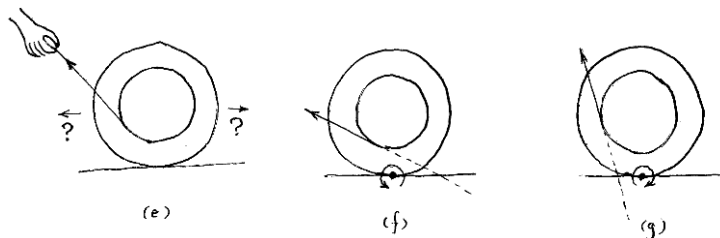
$$\Omega'R = \frac{5}{2} \frac{H - R}{R} V'$$

whence the velocity of the point of contact with the table is

$$V' - \Omega'R = \frac{7R - 5H}{2R} V'.$$

$H < \frac{7}{5}R$ induces backward spin, $H > \frac{7}{5}R$ forward spin, $H = \frac{7}{5}R$ pure rolling.

4.4. Gently tug on the string of a spool (picture e). Which way will the spool roll?



Two theories: 1) you input momentum in the direction of tugging, so the spool rolls left; 2) tugging induces clockwise spinning, so the spool rolls right.

Which way the spool rolls depends on the inclination of the tug. In picture (f), the line of force passes *above* the point of contact with the ground, so the tug creates *anticlockwise* angular momentum around the point of contact; the spool rolls left, reeling the string *in*. Likewise in picture (g), the spool rolls right, reeling the string *out*.

4.5. Place a ball on a sheet of paper, and withdraw the sheet from under the ball. Which way will the ball end up rolling? Two competing theories again. The answer is that the ball stops dead.

Moral of (4.4, 4.5): it can prove useful to consider angular momentum around points other than the centre of mass (2.8).

4.6. A *superball* is a perfectly elastic ball whose surface is non-slipping; *elastic* means no loss of energy upon bouncing, so a superball bounces excitingly high. We analyse the bouncing of a superball of mass M , radius R , moment of inertia $I = \frac{2}{5}MR^2$ around its centre (2.7).

Assume the problem is planar. The superball comes in at velocity whose *horizontal component* is V and angular velocity Ω around its centre, and bounces off a horizontal floor or ceiling; the vertical component of the velocity merely reverses upon bouncing. During the brief interval Δt of a bounce, the floor or ceiling exerts on the ball not only a normal reaction but also a friction F . Denoting the values after a bounce by $'$, we have from conservation of momentum and angular momentum (3.2)

$$M(V' - V) = F\Delta t, \quad I(\Omega' - \Omega) = -RF\Delta t$$

and from conservation of energy (3.3)

$$\frac{1}{2}MV'^2 + \frac{1}{2}I\Omega'^2 = \frac{1}{2}MV^2 + \frac{1}{2}I\Omega^2;$$

eliminating and factoring, we get two equations

$$M(V' - V) = -\frac{I}{R}(\Omega' - \Omega), \quad M(V' - V)(V' + V) = -I(\Omega' - \Omega)(\Omega' + \Omega).$$

The dull solution is $V' = V$, $\Omega' = \Omega$, $F\Delta t = 0$. The other solution, worthy of a superball, is

$$V' - \Omega'R = -(V - \Omega R), \quad F\Delta t = -\frac{2MI}{I + MR^2}(V - \Omega R),$$

i.e. upon bouncing the velocity of the point of contact instantaneously reverses: *a superball bounces not only normally but also tangentially*. The law of bouncing is then

$$\begin{aligned} V' &= \frac{3}{7}V + \frac{4R}{7}\Omega \\ \Omega' &= \frac{10}{7R}V - \frac{3}{7}\Omega \end{aligned}$$

for bounce off the floor, and

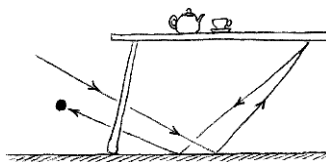
$$\begin{aligned} V' &= \frac{3}{7}V - \frac{4R}{7}\Omega \\ \Omega' &= -\frac{10}{7R}V - \frac{3}{7}\Omega \end{aligned}$$

for bounce off the ceiling. Both these linear operators have determinant -1 .

Throw a superball under a table (the underside of the table serving as ceiling). It bounces successively off: floor, ceiling, floor, ceiling...

$$\begin{aligned}
V & & V' &= \frac{3}{7}V & V'' &= -\frac{31}{49}V & V''' &= -\frac{333}{343}V \dots \\
\Omega &= 0 & \Omega' &= \frac{10}{7R}V & \Omega'' &= -\frac{60}{49R}V & \Omega''' &= -\frac{130}{343R}V \dots
\end{aligned}$$

The superball comes back out from under the table.



4.7. Lay a boiled egg, and give it a vigorous spin. It rises and spins upright (picture h). In fact, just about any convex object spun on a frictional surface tends to raise its centre of mass.

The simplest model of this phenomenon is as follows. To a hoop affix a wad of clay, and set it spinning about its diameter with the clay at the bottom. As the hoop spins, the clay rises to the top. In picture (i), the clay shifted the centre of mass C off the centre of curvature K of the hoop of radius R . The hoop plus the clay have mass M and a roughly spherical inertia matrix I around C . Gravity Mg presses the hoop down, provoking friction μMg at the point $\otimes$ directly beneath K . The angular momentum L around C is roughly vertical. In the configuration of picture (i), the spin plunges $\otimes$ into the page, so the friction protrudes *out* at $\otimes$. Its torque N around C is roughly horizontal. N makes L tremble, but because N whirls rapidly about L during the spin, L varies little on a long time scale—as observed in experiments.

We analyse the change in time of θ , the angle between L and the axis CK . For the component of L along CK (3.2),

$$\frac{d}{dt}|L|\cos\theta = -|N|\sin\theta.$$

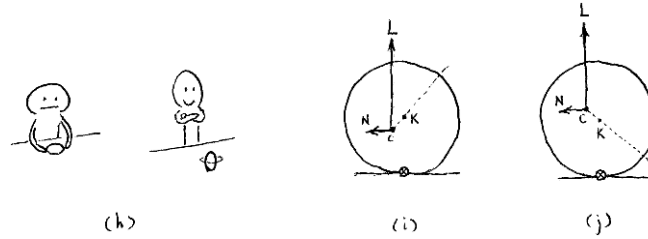
In the approximation of constant L

$$\frac{d\theta}{dt} \sim \frac{|N|}{|L|} \sim \frac{R\mu Mg}{I|\Omega|} > 0,$$

where in the same approximation Ω is the initial angular velocity given to the hoop. θ *increases*, which means CK *rises*. Lest readers worry what ensues once CK is horizontal, in picture (j) too θ goes on increasing; this shows incidentally that centrifugal force alone does not explain the phenomenon. The hoop tips over in time

$$\pi / \frac{d\theta}{dt} \sim \frac{\pi I |\Omega|}{R\mu Mg}.$$

For a commercially available **tippy top**, a wooden off-centered ball-like top, our theoretical value for the tip-over time is of the order of $\pi^{\frac{2}{5}} MR^2 |\Omega| / R\mu Mg \sim \pi^{\frac{2}{5}} \cdot 2 \text{ cm} \cdot 2\pi 50 \text{ Hz} / \frac{1}{3} \cdot 1000 \text{ cm sec}^{-2} \sim 2 \text{ sec}$.



The seemingly reckless approximations above are justifiable by a more precise analysis. For a physically important example, if you spin an egg too sluggishly, it rises only part of the way; the reason is that sliding at $\otimes$ transits to rolling and the friction coefficient μ drops. A precise analysis handles the sliding/rolling transition, among other things.

I also announce, for the first time in the literature, the existence of *chiral tippy tops*, which tip over when spun one way but not when spun the opposite way. They indicate that some crucial physical insight is missing from all previous theories of tippy top, none of which accommodates, let alone predicts, any chirality. I plan to publish a full discussion soon.

4.8. Too many books already treat gyroscopes.

4.9. How does a yo-yo work?

4.10. When leaves stop falling, fall starts leaving. Most falling leaves dance to and fro, zigzagging randomly earthbound. But there are some elongated leaves that spin busily about the long axis and fall along a fairly straight trajectory; the angular velocity is very large and roughly horizontal, the direction of the fall is roughly perpendicular to the angular velocity. Ditto for rectangular strips of paper: beyond a certain aspect ratio of the rectangle, they “tumble rather than flutter”. Why?

5 Euler’s Equations

5.1. A top T of inertia matrix I around a point O is spinning at angular velocity Ω around O . Let e_1, e_2, e_3 be the orthonormal basis vectors that define coordinates x_1, x_2, x_3 attached to T whose origin is at O . For any vector-valued function $A = A(t) = A_1 e_1 + A_2 e_2 + A_3 e_3$,

$$\frac{d}{dt} A = \left(\frac{dA_1}{dt} e_1 + \frac{dA_2}{dt} e_2 + \frac{dA_3}{dt} e_3 \right) + \left(A_1 \frac{de_1}{dt} + A_2 \frac{de_2}{dt} + A_3 \frac{de_3}{dt} \right).$$

We shall denote the first $(\dots)$ by $\frac{\partial}{\partial t}A$; on account of $\frac{d}{dt}e_i = \Omega \wedge e_i$, the second $(\dots)$ is $\Omega \wedge A$. Symbolically,

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \Omega \wedge .$$

Writing out $\frac{d}{dt}L = \frac{\partial}{\partial t}L + \Omega \wedge L = I \frac{\partial}{\partial t}\Omega + \Omega \wedge I\Omega = N$ (3.2) with respect to principal axes, we obtain **Euler's equations** [5]

$$\begin{aligned} I_1 \frac{\partial}{\partial t}\Omega_1 &= (I_2 - I_3)\Omega_2\Omega_3 + N_1 \\ I_2 \frac{\partial}{\partial t}\Omega_2 &= (I_3 - I_1)\Omega_3\Omega_1 + N_2 \\ I_3 \frac{\partial}{\partial t}\Omega_3 &= (I_1 - I_2)\Omega_1\Omega_2 + N_3, \end{aligned}$$

the torque N being around O . Though something of an elephant in a china shop when applied to concrete problems, Euler's equations are effective in theoretical investigations: cf. Sects. 6, 7, 8.

5.2. Euler's equations in hydrodynamics for an ideal fluid are interpretable as Euler's equations for an infinite-dimensional rigid body [2].

5.3. Essentially three kinds of tops have been studied in the literature:

- Euler's top
- Lagrange's top
- Kovalevskaya's top.

Moreover, it is a theorem that these tops and these alone are *algebraically integrable*. We shall study them in turn: Euler in Sect. 6, Lagrange in Sect. 7, Kovalevskaya in Sect. 8.

6 Spinning under No Torque: Euler's Top

6.1. Throughout this section, the force and the torque are absent

$$F = 0, \quad N = 0,$$

which implies constant momentum, angular momentum, energy; modulo a Galilean transformation we may even assume that P is zero:

$$P = 0, \quad L = I\Omega = \text{const.}, \quad E = E_{\text{rot}} = \text{const.}$$

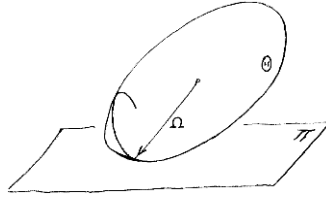
Such a rigid body, in “free rotation” around its immobile centre of mass, is called **Euler's top** [6]. Isolated celestial bodies are examples, as are gyroscopes supported at their centres of mass. We describe the motion of Euler's top in two ways: pictorial (6.2, 6.3) and analytical (6.4, 6.5, 6.6, 6.7).

6.2. Poincot [13] devised a pictorial description of Euler's top. The ingredients of the picture are built from the constants of the top: matrix I , scalar E , vector L . The description revolves around the distinction between L and Ω (2.2): L is constant but in general Ω moves.

Imagine an ellipsoid *attached to the top*

$$\Theta : \langle Ix, x \rangle = 2E$$

and a plane *fixed in space*



$$\Pi : \langle L, x \rangle = 2E.$$

The trick now is to consider the point $x = \Omega$ of Θ . On one hand, the tangent plane to Θ at $x = \Omega$ is Π (its equation being $2\langle I\Omega, x \rangle = 2\langle I\Omega, \Omega \rangle$). On the other hand, since the top is instantaneously spinning about Ω , $x = \Omega$ is instantaneously at rest. These together mean that

Euler's top moves as if the ellipsoid Θ were rolling on the plane Π .

The curve traced on Θ [resp. Π] by the point of rolling contact $x = \Omega$ is the **polhode** [resp. **herpolhode**]. In principle the motion of the top can be reconstructed from the polhode.

6.3. With respect to principal axes

$$E = \frac{1}{2}(I_1\Omega_1^2 + I_2\Omega_2^2 + I_3\Omega_3^2), \quad L^2 = I_1^2\Omega_1^2 + I_2^2\Omega_2^2 + I_3^2\Omega_3^2,$$

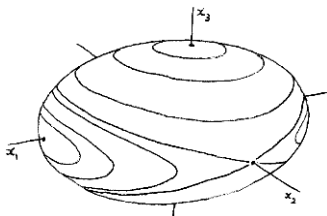
so

$$\text{polhode} = \{E = \text{const.}\} \cap \{L^2 = \text{const.}\}.$$

Switching to the variables L_1, L_2, L_3 facilitates visualisation:

$$\text{polhode} = \left\{ \frac{L_1^2}{2EI_1} + \frac{L_2^2}{2EI_2} + \frac{L_3^2}{2EI_3} = 1 \right\} \cap \left\{ L_1^2 + L_2^2 + L_3^2 = L^2 \text{ (const.)} \right\},$$

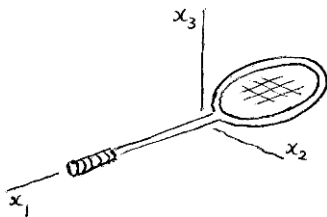
i.e. a polhode is a curve along which an ellipsoid and a sphere intersect. As various values of L and E are picked, a family of such curves are cut out. The choice of an initial condition puts Ω on one of these curves, and from then on Ω follows that curve.



The picture has been drawn assuming $I_1 < I_2 < I_3$. It shows that a polhode starting near the x_3 - or x_1 -axis dawdles near that axis, whereas a polhode starting near the x_2 -axis wanders far from that axis and swings over to the other side of the ellipsoid.

Suppose $I_1 < I_2 < I_3$. Then the rotation of the top is stable about x_3 and x_1 , unstable about x_2 .

This stability result is nicknamed “tennis racket theorem”: a racket tossed spinning is easy to catch if spun about x_3 or x_1 , but it wobbles out of control if spun about x_2 .



Poinsot’s picture tells us the *trajectory* of Euler’s top. What it leaves untold is at what pace the top follows the trajectory *in the course of time*. The time-evolution is rendered explicit by the analytical description. We analyse cases of increasing generality.

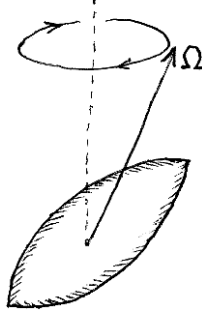
6.4. Case of a *spherical* top, $I_1 = I_2 = I_3$. Euler’s equations (5.1) reduce to $\frac{\partial}{\partial t}\Omega = 0$, $\Omega = \text{const.}$: the top continues to spin about the same axis at the same rate—quite uneventful.

6.5. Case of a *symmetric* top, $I_1 = I_2 \neq I_3$ —slightly more eventful. Euler’s equations (5.1) may be recast as

$$\frac{\partial}{\partial t}(\Omega_1 + i\Omega_2) = i\Omega_3\left(\frac{I_3}{I_1} - 1\right)(\Omega_1 + i\Omega_2), \quad \frac{\partial}{\partial t}\Omega_3 = 0,$$

which integrate to

$$\Omega_1 + i\Omega_2 = (\Omega_1(0) + i\Omega_2(0)) \exp\left[i\Omega_3(0)\left(\frac{I_3}{I_1} - 1\right)t\right], \quad \Omega_3 = \Omega_3(0) \quad (\text{const.}).$$



The top **precesses** with period $2\pi/|\Omega_3(0)(I_3/I_1 - 1)|$; the period does not depend on Ω_1, Ω_2 , i.e. not on how widely Ω is tilted away from $(0, 0, \Omega_3)$.

Since the oblateness of the Earth (extra bulge at the equator) is $I_3/I_1 \sim 301/300$, and $\Omega_3 = 2\pi/1$ day, our theoretical value for the precession period of the Earth is ~ 300 days. The observed value, the “Chandler period”, is ~ 440 days.

The limit $I_3 \rightarrow I_1$ yields $I_3/I_1 - 1 \rightarrow 0$, trigonometric functions degenerate to constants, recovering the spherical case (6.4).

6.6. Generic case of Euler’s top. It turns out the problem is integrable in terms of Jacobian elliptic functions [8] (reference on elliptic functions: [9]).

Recall the conservation laws

$$E = \frac{1}{2}(I_1\Omega_1^2 + I_2\Omega_2^2 + I_3\Omega_3^2), \quad L^2 = I_1^2\Omega_1^2 + I_2^2\Omega_2^2 + I_3^2\Omega_3^2.$$

The principal moments of inertia are all distinct, say $I_1 < I_2 < I_3$. Then $I_1 < L^2/2E < I_3$. In the picture (6.3), the separatrices slice the ellipsoid into 4 eye-shaped sectors $\Omega_3 > 0$, $\Omega_3 < 0$ and $\Omega_1 > 0$, $\Omega_1 < 0$, the former two satisfying $L^2/2E > I_2$ and the latter two $L^2/2E < I_2$. Let us analyse a motion during which Ω_3 keeps a constant sign (for Ω_1 constant sign swap the indices 3 and 1). Extracting Ω_3^2, Ω_1^2 between the conservation laws,

$$\Omega_3^2 = \frac{L^2 - 2EI_1 - (I_2 - I_1)I_2\Omega_2^2}{(I_3 - I_1)I_3}, \quad \Omega_1^2 = \frac{L^2 - 2EI_3 - (I_2 - I_3)I_2\Omega_2^2}{(I_1 - I_3)I_1},$$

which separate the second of Euler’s equations (5.1)

$$\frac{\partial}{\partial t}\Omega_2 = \frac{I_3 - I_1}{I_2}\Omega_3\Omega_1 = \sqrt{\text{polynomial of degree 4 in } \Omega_2}.$$

In rescaled variables

$$\tau = t\sqrt{\frac{(I_3 - I_2)(L^2 - 2EI_1)}{I_1I_2I_3}}, \quad \omega = \Omega_2\sqrt{\frac{(I_2 - I_3)I_2}{L^2 - 2EI_3}}$$

and a new constant (*modulus*)

$$k^2 = \frac{(I_1 - I_2)(L^2 - 2EI_3)}{(I_3 - I_2)(L^2 - 2EI_1)} \quad (0 < k^2 < 1),$$

the equation $\frac{\partial}{\partial t}\Omega_2 = \dots$ integrates to

$$\tau = \int_0^\omega \frac{d\omega}{\sqrt{(1-\omega^2)(1-k^2\omega^2)}}.$$

Inverting, we find $\omega = \operatorname{sn} \tau$, which as a function of t determines Ω_2 and thereby Ω_3, Ω_1 :

$$\Omega_1 = \sqrt{\frac{L^2 - 2EI_3}{(I_1 - I_3)I_1}} \operatorname{cn} \tau, \quad \Omega_2 = \sqrt{\frac{L^2 - 2EI_3}{(I_2 - I_3)I_2}} \operatorname{sn} \tau, \quad \Omega_3 = \sqrt{\frac{L^2 - 2EI_1}{(I_3 - I_1)I_3}} \operatorname{dn} \tau.$$

The period in t is

$$4K(k) \sqrt{\frac{I_1 I_2 I_3}{(I_3 - I_2)(L^2 - 2EI_1)}}.$$

The limit $I_2 \rightarrow I_1$ yields $k^2 \rightarrow 0$, elliptic functions degenerate to trigonometric ones, recovering the symmetric case (6.5).

6.7. Tennis racket revisited. Earlier the stability result (6.3) was deduced pictorially. Analytically it could be excavated from the exact solution (6.6). More cheaply, perturb $\Omega = (0, \Omega_2(0), 0)$, a steady rotation about x_2 , to $(\Delta\Omega_1, \Omega_2(0) + \Delta\Omega_2, \Delta\Omega_3)$. Neglecting terms of order Δ^2 or higher in Euler's equations (5.1),

$$\frac{\partial}{\partial t}\Omega_2 = 0, \quad \frac{\partial^2}{\partial t^2}\Delta\Omega_i = \lambda\Delta\Omega_i \quad (i = 3, 1) \quad \text{with } \lambda = (I_1 - I_2)(I_2 - I_3)/I_3 I_1 > 0.$$

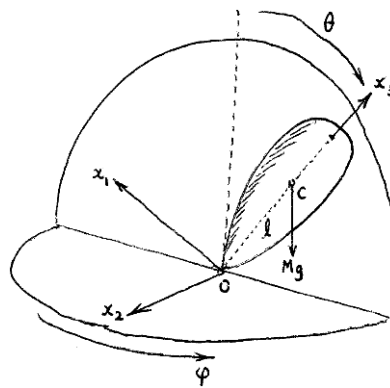
Unless the perturbation puts Ω on an incoming separatrix in Poincaré's picture (6.3), $\Delta\Omega_i$ contains an exponential term with exponent $+\sqrt{\lambda} > 0$, so rotation about x_2 is *unstable*. Similarly rotation about x_3 or x_1 is *stable*.

6.8. It is no accident that integrable problems involve elliptic—or rather theta—functions, for geometrically integrability means foliation of the phase space into invariant tori, and theta functions are the very creatures, via Abel-Jacobi embeddings, that give us holomorphic functions on a torus. But I digress.

7 Some Cases of Spinning under Torques: Lagrange's Top

7.1. This section studies a top friendlier than Euler's but in a more hostile environment: **Lagrange's top** [12] is symmetric, $I_1 = I_2$, pivoted at a point *on the axis of symmetry but not at the centre of mass* and spinning under

7.2. Lagrange's top T of mass M is spinning, tilted at an angle θ (colatitude) from the vertical. T swings about the vertical by an angle φ (longitude). Let ℓ be the distance from the pivot O to the centre of mass C of T . At the instant under consideration, take x_3 along the top's axis of symmetry, x_2 horizontal and perpendicular to x_3 , x_1 perpendicular to the x_2x_3 -plane, the axes having their origin at O . The inertia matrix I is around O , not around C .



$$L_{\text{vert}} = I_1 \Omega_1 \sin \theta + I_3 \Omega_3 \cos \theta = I_1 \frac{d\varphi}{dt} \sin^2 \theta + L_3 \cos \theta = \text{const.}$$

10

$$= \frac{1}{2} I_1 \left[\left(\frac{d\theta}{dt} \right)^2 + \left(\frac{d\varphi}{dt} \right)^2 \sin^2 \theta \right] + \frac{L_3^2}{2I_3} + M g \ell \cos \theta = \text{const.}$$

$$h = \cos \theta$$

○

$$I_1^2 \left(\frac{dh}{dt} \right)^2 = 2I_1 \left(E - \frac{L_3^2}{2I_3} - Mglh \right) (1 - h^2) - (L_{\text{vert}} - L_3 h)^2.$$

The right-hand side $f(h)$ is a cubic polynomial in h , with roots say h_1, h_2, h_3 . The equation integrates to

$$h = h_1 + (h_2 - h_1) \operatorname{sn}^2 \left(t \sqrt{\frac{Mgl(h_3 - h_1)}{2I_1}} \right)$$

with modulus

$$k^2 = \frac{h_2 - h_1}{h_3 - h_1}.$$

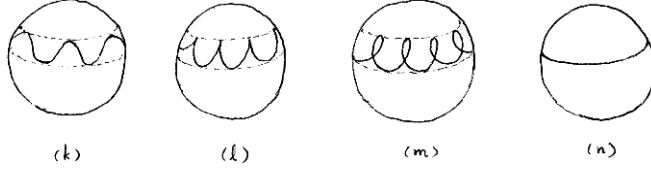
This determines h , thereby θ , as a periodic function of t , **nutatation**; its period is

$$2K(k) \sqrt{\frac{2I_1}{Mgl(h_3 - h_1)}}.$$

In its turn, φ is determined as an elliptic integral

$$\varphi = \int_0^h \frac{L_{\text{vert}} - L_3 h}{(1 - h^2) \sqrt{f(h)}} dh.$$

Generically the axis of symmetry of T traces waves (picture k) or swirls (picture m).



7.3. In *pure precessions*, i.e. precessions with zero nutation (picture n), θ , or h , is constant, so $h_2 - h_1 = 0$. Therefore pure precessions are sustained at a tilt angle $\theta_{\text{pr}} = \arccos h_{\text{pr}}$ that satisfies the double-root condition

$$f(h_{\text{pr}}) = f'(h_{\text{pr}}) = 0.$$

Combining this with the conservation laws leads to

$$I_1 \cos \theta_{\text{pr}} \left(\frac{d\varphi_{\text{pr}}}{dt} \right)^2 - L_3 \frac{d\varphi_{\text{pr}}}{dt} + Mgl = 0,$$

an equation quadratic in the rate of pure precession $d\varphi_{\text{pr}}/dt$. Suppose that “spin overwhelms gravity”: $L_3^2 \gg I_1 \cos \theta_{\text{pr}} Mgl$. The binomial expansion of the roots then yields

$$\text{slow precession} \quad \frac{d\varphi_{\text{pr}}}{dt} \sim \frac{Mg\ell}{L_3}, \quad \text{fast precession} \quad \frac{d\varphi_{\text{pr}}}{dt} \sim \frac{L_3}{I_1 \cos \theta_{\text{pr}}}.$$

The fast precession tends to be damped away quickly.

7.4. If a spinning top is released from a tilted position, it *dips* at first, then goes into precession and nutation (picture 1). The graph of θ against φ is approximately a cycloid. In a real top, as friction at the pivot damps the nutation, the motion asymptotes to a pure precession.

7.5. If a spinning top is released upright, $\theta = 0$, $h = 1$, it may be able to stay upright; this is the **sleeping top**. A sleeping top is stable provided $f(h) < 0$ near $h = 1$, i.e.

$$\Omega^2 = \Omega_3^2 > \frac{4I_1 Mg\ell}{I_3^2}.$$

So a top needs to be spun sufficiently fast to go to sleep. In a real top, friction decelerates Ω ; when eventually Ω violates the above inequality, the top *wakes up* and goes into precession and nutation. Conversely, if a top is spun sufficiently fast, even from a tilted position it snaps upright and goes to sleep, by the tippy-top mechanism (4.7).

7.6. In the limit $I_3 \rightarrow 0$, Lagrange's top degenerates to a spherical pendulum. As a corollary a spherical pendulum is integrable in terms of elliptic functions.

8 Kovalevskaya's Top

8.1. Our final top T also spins under gravity. As with Lagrange's top (7.1), pivot T at a point O not its centre of mass C and take x_1, x_2, x_3 principal axes attached to T with their origin at O . The inertia matrix I is around O . Let (C_1, C_2, C_3) be the (constant) coordinates of the centre of mass, (z_1, z_2, z_3) the (variable) components of the upward unit vector z . Euler's equations (5.1) are

$$\begin{aligned} I_1 \frac{\partial}{\partial t} \Omega_1 &= (I_2 - I_3) \Omega_2 \Omega_3 - Mg(C_2 z_3 - C_3 z_2) \\ I_2 \frac{\partial}{\partial t} \Omega_2 &= (I_3 - I_1) \Omega_3 \Omega_1 - Mg(C_3 z_1 - C_1 z_3) \\ I_3 \frac{\partial}{\partial t} \Omega_3 &= (I_1 - I_2) \Omega_1 \Omega_2 - Mg(C_1 z_2 - C_2 z_1). \end{aligned}$$

T has 3 degrees of freedom and 2 conserved quantities E, L_{vert} (7.2). In comparison with Lagrange's top, we *lose the conserved quantity* L_3 because we are no longer assuming that OC is an axis of symmetry of T . In order to integrate the problem, we need 1 more conserved quantity. **Kovalevskaya's top** [11] is exactly rigged so as to allow the existence of a third conserved quantity.

*Kovalevskaya's top: $I_1 = I_2 = 2I_3$
and the centre of mass C is on the x_1x_2 -plane.*

E.g. a homogeneous ellipsoid with semiaxes $1, \sqrt{3}, 3$ pivoted on the x_1 -axis at a distance $\sqrt{2/5}$ from the centre.

8.2. Without loss of generality set $I_1 = I_2 = 2, I_3 = 1, C_2 = C_3 = 0$. Euler's equations (8.1) become

$$\begin{aligned} 2\frac{\partial}{\partial t}\Omega_1 &= \Omega_2\Omega_3 \\ 2\frac{\partial}{\partial t}\Omega_2 &= -\Omega_3\Omega_1 + MgC_1z_3 \\ \frac{\partial}{\partial t}\Omega_3 &= -MgC_1z_2. \end{aligned}$$

Writing out $0 = \frac{d}{dt}z = \frac{\partial}{\partial t}z + \Omega \wedge z$ (5.1) in coordinates,

$$\frac{\partial}{\partial t}z_1 = z_2\Omega_3 - z_3\Omega_2, \quad \frac{\partial}{\partial t}z_2 = z_3\Omega_1 - z_1\Omega_3, \quad \frac{\partial}{\partial t}z_3 = z_1\Omega_2 - z_2\Omega_1.$$

Claim:

Kovalevskaya's top has the conserved quantity $|(\Omega_1 + i\Omega_2)^2 - MgC_1(z_1 + iz_2)|$.

Indeed,

$$\begin{aligned} 2\frac{\partial}{\partial t}(\Omega_1 + i\Omega_2) &= -i[(\Omega_1 + i\Omega_2)\Omega_3 - MgC_1z_3], \\ \frac{\partial}{\partial t}(z_1 + iz_2) &= -i[(z_1 + iz_2)\Omega_3 - z_3(\Omega_1 + i\Omega_2)], \end{aligned}$$

therefore

$$\frac{\partial}{\partial t}\{(\Omega_1 + i\Omega_2)^2 - MgC_1(z_1 + iz_2)\} = -i\Omega_3\{(\Omega_1 + i\Omega_2)^2 - MgC_1(z_1 + iz_2)\}.$$

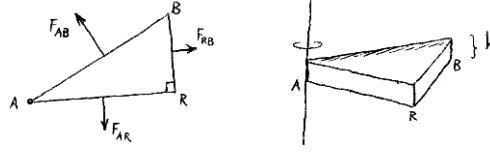
Since the velocity of $\{\dots\}$ is perpendicular to $\{\dots\}$, the absolute value $|\{\dots\}| = \text{const.}$

The integration is completed in terms of hyperelliptic functions [7]. In the limit $C_1 \rightarrow 0$, Kovalevskaya's top degenerates to a special case of Lagrange's top.

8.3. Kovalevskaya's top was the last integrable system of the 19th century. The discovery of the next integrable system had to wait 78 years, until Toda lattices arrived on the scene [16].

9 Appendix

9.1. Let ARB be a right triangle. We wish to prove that $AR^2 + RB^2 = AB^2$. Upon ARB as base build a box of height h and hinge it at A to a vertical axis, around which it can revolve smoothly.



Now fill the box with gas of pressure p . The gas exerts forces that may be regarded as acting at the centre of, and normal to, each face of the box. The forces on the lid and the bottom don't interest us. The forces F_{AR}, F_{RB} on the sides AR, RB try to revolve the box clockwise, whereas the force F_{AB} on the side AB tries to revolve it anticlockwise. But filling with gas can't coax a box into moving: the torques about the axis must balance. The torques due to F_{AR}, F_{AB} are $AR/2 \times F_{AR}, AB/2 \times F_{AB}$, and *because R is a right angle* the torque due to F_{RB} is $RB/2 \times F_{RB}$:

$$\frac{AR}{2} \times F_{AR} + \frac{RB}{2} \times F_{RB} = \frac{AB}{2} \times F_{AB}.$$

Force is pressure times area, $F_{AR} = phAR$, etc. Dividing through by $ph/2$, we are home.

9.2. Recycling the argument on a not necessarily right triangle proves the “cosine law”.

10 Further Reading and Acknowledgement

Dynamics of rigid bodies in rotation is a staple diet of textbooks on mechanics [1]. Among specialised monographs, the richest cache of examples is [14, 15]. *μέγα βιβλίου μέγα κακόν* to [10], though admittedly it makes available material not collected elsewhere. [4] is elementary and charming; inevitably for elementary charming books, it is out of print. [7] exposes the relationship between spinning tops and elliptic/theta functions. To acquaint yourself with the current mathematical take on the subject, [3].

This article reproduces lectures, minus toy demonstrations, from the CNRS *école d'hiver* at Lanslevillard, March 2003. I thank its organiser J. Souchay for his kind invitation and J. Laskar for first suggesting that these lectures be given. I am also obliged to R. and D. Gonczi for their hospitality in Nice.

References

1. P. Appell, *Traité de Mécanique Rationnelle*, tome II, Gauthier-Villars, 1896; repr. J. Gabay, 1991.
2. V. Arnold, Ann. Inst. Fourier, **16** (1966) 319–361.

3. M. Audin, *Spinning Tops*, Cambridge UP, 1996.
4. H. Crabtree, *Elementary Treatment of the Theory of Spinning Tops and Gyroscopic Motion*, Longmans, 1909; repr. Chelsea, 1967.
5. L. Euler, Mém. de l'Acad. Sci. Berlin **14** (1758) 154–193 $\simeq$ *Opera* II **7**, 200–235.
6. L. Euler, *Theoria Motus Corporum Solidorum seu Rigidorum*..., A. F. Roese, 1765 $\simeq$ *Opera* II **3**, **4**.
7. V. V. Golubev, *Lektsii po Integrirvaniyu Uravnenii Dvizheniya Tyazhelogo Tverdogo Tela okolo Nepodvizhnoi Tochki*, Gostekhizdat, 1953.
8. C. G. J. Jacobi, Crelle **39** (1849) 293–350 $\simeq$ *Werke* **2**, 289–352.
9. C. Jordan, *Cours d'Analyse*, tome II, Gauthier-Villars, 2e ed. 1894; repr. J. Gabay, 1991.
10. F. Klein & A. Sommerfeld, *Über die Theorie des Kreisels*, Band I–IV, Teubner, 1897–1910.
11. S. Kowalevski, Acta Math. **12** (1889) 177–232.
12. J.-L. Lagrange, *Mécanique Analytique*, Veuve Desaint, 1788 $\simeq$ *Œuvres* **11**, **12**; repr. J. Gabay, 1989.
13. L. Poinsot, *Théorie Nouvelle de la Rotation des Corps*, Paris, 1834; repr. J. Math. Pures Appl., série I, **16** (1851) 9–129, 289–336.
14. E. J. Routh, *Dynamics of a System of Rigid Bodies: Elementary Part*, Macmillan, 1860; repr. Dover, 1960.
15. E. J. Routh, *Dynamics of a System of Rigid Bodies: Advanced Part*, Macmillan, 1860; repr. Dover, 1955.
16. M. Toda, J. Phys. Soc. Japan **22** (1967) 431–436.

Physics Inside the Earth: Deformation and Rotation

Hilaire Legros¹, Marianne Greff², and Tadashi Tokieda³

¹ E.O.S.T., 5 rue René Descartes, 67084 Strasbourg-Cedex, France
Hilaire.Legros@eost.u-strasbg.fr

² I.P.G.P., 4 place Jussieu, 75252 Paris-Cedex 05, France
greff@ipgp.jussieu.fr

³ Trinity Hall, Cambridge CB2 1TJ, UK
tokieda@dpms.cam.ac.uk

1 Introduction

This paper presents a theoretical basis for elastic and viscoelastic deformations of the Earth and analyses the rotation of a deformable planet having a fluid core and a solid inner core. Section 2 reviews the concepts that become indispensable when we pass from the model of the Earth as a rigid body to its more realistic model as a deformable planet. In this realistic model, variations in the physical parameters allow us to understand convection within the mantle, stratification into solid and fluid parts, and evolution of the Earth's density and inertia tensor. Section 3 addresses detailed problems in the theory of deformations, in particular the effect of rheology. Finally Sect. 4 studies how a deformable, stratified Earth actually rotates.

2 Terrestrial Mechanics and Survey of Some Dynamical Theories

2.1 Historical Review

The Earth is a planet of the solar system. But what does 'planet' mean, beyond a pretty picture of a spherical body with a topography revolving around the Sun? To make this term precise, let us recall the key steps that have led to our present understanding of the Earth. A primitive, but already fairly accurate, view of the Earth was that of a stratified medium with a fluid core in which mechanical parameters such as density or elastic modulus varied as functions of depth. This model was built in 1936–1942 by Bullen, Jeffreys, and Birch (cf. [16]). By clocking how long seismic waves took to travel from an earthquake to an observing station and using the Herglotz-Wiechert inversion formula (1909–1910), one could deduce the distribution of the seismic

velocities of P and S waves as functions of depth, and discover internal discontinuities. From this and the Adams-Williamson relation (1923), Bullen proposed in 1936 a first model of the Earth, which underwent subsequent improvements until 1942. The 1942 model consisted of a 30 km-thick crust, a 1st layer down to 400 km deep followed by a 2nd layer down to 700 km deep, then a 3rd layer down to 3000 km deep [26]. The temperature distribution was unknown and the composition was assumed to come from a rock named peridotite which may be modified to give the discontinuities. Below 3000 km sat the core, thought to be fluid [35] and composed of iron and nickel [67]. A solid inner core occupied the center of the Earth [40]. This is the seismologically inspired model; combined with geodetical measurements of the ratio C/Ma it has eventually led to the reference model PREM [17].

A new line of attack on the problem of the Earth's interior started in 1960–1970, inspired by different works:

- Discovery of surface dynamics [29], culminating in the theory of plate tectonics [41, 45].
- Researches on the chemical composition, and on the mineralogical phase changes depending on pressure and temperature, became integrated with researches on the chemical composition of the Sun and meteorites as well as the theory of planetary accretion [61]. The age of 4.50 Ga for the solar system had been proposed by Patterson (1956).
- 1966 brought the first picture of terrestrial gravity field collated from artificial satellites. The lunar mission Apollo 11 reinforced the image of geophysics as a science of planetary caliber.
- Based on these works, the present frame of thinking emerged in 1970–1985:
 - perovskite structure for the lower mantle [42]
 - first thermal models [3, 66]
 - theories of accretion [9, 72] revealing a hot origin of the planets and allowing the first thermal histories to be put together
 - dynamical theories of deformation and rotation (e.g. Sasao et al. 1980)
 - viscoelastic theory and post-glacial rebound [57]

From 1985, models of planetary convection began appearing, building on earlier more theoretical studies. It is also from this period that a major development of thermodynamical studies on phase transitions started.

This historical review was sketchy in the extreme. Nevertheless it underscores the youth of these dynamical theories which are still in full growth spurt. It is in the light of these growing theories that we offer the presentation that follows.

2.2 Physical and Mechanical Setup

For further details on this subsection, cf. [4, 11, 58, 64].

In the eyes of a physicist, the Earth – as a telluric planet – is a large quasi-spherical body about 6400 km in radius. Write V for its volume. A

volume element in this environment comes with parameters that vary during the spatio-temporal evolution of V :

- position, characterized by its velocity $\mathbf{v}$
- thermodynamic state, characterized by:
 - deformation (displacement $\mathbf{u}$, strain tensor ϵ_{ij} , strain rate as a function of time $d\epsilon_{ij}/dt = d_{ij}$), which causes variations in density ρ
 - entropy S
 - pressure P and more generally stress tensor σ_{ij}
 - temperature T
- phase of the material in this volume
- diffusion across the boundary, in particular heat flux $\mathbf{q}$ per unit time and surface area

The variations of these parameters are caused by body forces $\rho\mathbf{f}$, surface forces, and the rate of heat production r .

The same volume element also comes with parameters of its chemical and mineralogical structure:

- density ρ
- calorimetric parameters: thermal expansion $\alpha = -(\partial\rho/\partial T)_P/\rho$, specific heat at constant pressure $C_p = T(\partial S/\partial T)_P$
- elastic parameters: modulus of incompressibility at constant temperature $K_T = \rho(\partial P/\partial\rho)_T$, shear modulus $\mu = \frac{1}{2}\partial\sigma_{ij}/\partial\epsilon_{ij}$ for $i \neq j$, isentropic modulus of incompressibility $K_S = \rho(\partial P/\partial\rho)_S$, specific heat at constant volume $C_v = T(\partial S/\partial T)_V$
- Grüneisen parameter: $\gamma = \alpha K_S/\rho C_p = \alpha K_T/\rho C_v$

The diffusion processes are governed by parameters such as thermal conductivity κ and viscosity η . We do not take into account phase changes and chemical transformations.

How does this volumetric element evolve? Mechanics dictates conservations of mass and momentum:

$$\begin{aligned}\partial\rho/\partial t + \nabla \cdot (\rho\mathbf{v}) &= 0 \\ \rho d\mathbf{v}/dt &= \nabla \cdot \boldsymbol{\sigma} + \rho\mathbf{f}\end{aligned}\tag{1}$$

Thermodynamic of reversible processes dictate:

$$\begin{aligned}\rho dU/dt &= \sigma_{ij}d_{ij} - \nabla \cdot \mathbf{q} + r \\ \rho T dS/dt &= -\nabla \cdot \mathbf{q} + r\end{aligned}\tag{2}$$

where U and S are the internal specific energy and the specific entropy. Thermodynamics gives further equations. In the absence of phase transition, we have 2 differential equations and the conservation of energy ('heat equation')

$$\begin{aligned}
\rho dS &= \rho C_p dT/T - \alpha dP \\
d\rho/\rho &= dP/K_T - \alpha dT = dP/K_S - \alpha T dS/C_p \\
\rho C_p dT/dt - \alpha T dP/dt &= -\nabla \cdot \mathbf{q} + r
\end{aligned} \tag{3}$$

We are nearly done. The equations of rheology or of transport read

$$\begin{aligned}
\mathbf{q} &= -\kappa \nabla T \\
\sigma_{ij} &= (K - \frac{2}{3}\mu)(\nabla \cdot \mathbf{u})\delta_{ij} + 2\mu\epsilon_{ij} \quad \text{for elastic body} \\
\sigma_{ij} &= -\frac{2}{3}\eta(\nabla \cdot \mathbf{v})\delta_{ij} + 2\eta d_{ij} \quad \text{for Newtonian fluid}
\end{aligned} \tag{4}$$

$K - \frac{2}{3}\mu$ is usually denoted by λ , Lamé parameter; here $K = K_T$ or K_S accordingly as the process is isothermal or isentropic (cf. 2nd equation in (3)). Finally the equations of state, in the form $f(P, V, T) = 0$ or in any of the partial forms $\alpha(P, T)$, $K_P(P, T)$, $K_T(P, T)$, $\gamma(P, T)$, $\eta(P, T)$, $\kappa(P, T)$, $f(P, V) = 0$ connect variations of these parameters.

To complete this setup of equations, we have to add the equations for the gravitational potential Φ :

$$\begin{aligned}
\Delta\Phi &= -4\pi G\rho \quad \text{Poisson's equation} \\
\mathbf{f} &= \nabla\Phi
\end{aligned} \tag{5}$$

As you can see, our setup neglects electromagnetism.

The analysis of these equations provides information on:

- evolution of the parameters
- evolution of the form of the Earth
- convection inside the Earth
- variations in density and inertia tensor. This last information is crucial in the study of rotation.

2.3 Classical Theories

It is one thing to have a full set of equations, another thing to extract from them a global theory of terrestrial dynamics. The problem is too complicated: the equations of state and of rheological behavior are only imperfectly known, so comparisons of their integrals with observations will not be convincing. The sensible policy is to analyse a subset of these equations, to develop partial theories, whose results can be easily compared with observations. We shall now survey a few such partial theories just to show the role they can play in the theory of deformation and rotation.

But first a word to the wise. Understanding the Earth involves more than solving equations. There is, most importantly, feedback from seismological data. Via recordings of seismological stations we can sound the Earth's interior very precisely. We thus know that the Earth is stratified into an elastic mantle composed of silicates, a fluid core composed essentially of iron, and a solid inner core composed of iron. We know too the distribution of density and

elastic parameters as functions of depth. Moreover, tomographic studies have revealed lateral variations in density, associated with how hot the material is. Besides seismological data, many geological, geophysical, and geodetical observations show that the Earth is deformed by tides and surface loading, that its external shell is made of lithospheric plates sliding one atop another, and that surface deformations occur on a time-scale ranging from 10^3 to 10^6 years. Thanks to all these sciences we have a knowledge of the planet Earth incomparably better than of other planets. We should keep that in mind while wending through partial theories of terrestrial dynamics.

Hydrostatic Figure of the Earth

A first partial theory, simple but important, is the theory of hydrostatic figure of the Earth (e.g. [50]). The results of this theory are in good agreement with observations. We keep just the conservation of momentum and Poisson's equation:

$$\begin{aligned} -\nabla P + \rho \nabla(\Phi + \Psi) &= \mathbf{0} \\ \Delta \Phi &= -4\pi G \rho \end{aligned} \tag{6}$$

where the new potential Ψ represents rotation of the planet. ρ is assumed known or given by the equation of state $P(\rho) = 0$. 2 important consequences of this theory are:

- Good explanation of the figure of the Earth due to the combined effects of its rotation and gravitation. Because $\Psi \ll \Phi$, the planets are nearly spherical. Let I be the average $I = \frac{1}{3}(A + B + C)$ of the principal moments of inertia A, B, C . The deviation of A, B, C , caused by the rotation, is weak and can be treated as a perturbation, a fundamental point in the study of rotation.
- The rheological behavior of the planet on the time-scale of the age of the Earth is quasi-fluid. This quasi-fluid behavior produces a fairly good explanation of the isostatic equilibrium of the continental and oceanic crust.

Convection

A second partial theory, more complicated than the first, is that of convection (e.g. [56]). This time we keep the conservation of mass and momentum, and Poisson's equation:

$$\begin{aligned} \nabla \cdot (\rho \mathbf{v}) &= 0 \\ \rho d\mathbf{v}/dt &= -\nabla P + \nabla \cdot \sigma + \rho \nabla \Phi \quad \text{with} \quad \sigma_{ij} = 2\eta d_{ij} \\ \Delta \Phi &= -4\pi G \rho \end{aligned} \tag{7}$$

to which we adjoin the conservation of energy (or of heat flux):

$$\rho C_P \frac{dT}{dt} - \alpha T \frac{dP}{dt} = \nabla \cdot (\kappa \nabla T) + r \quad (8)$$

The adiabatic variation of temperature (dT_a) says:

$$\frac{dT_a}{T_a} = \frac{\gamma}{K_S} dP \quad (9)$$

The equation of state connects α , C_P , γ and sometimes η . Finally, the perturbation of density is given by

$$\frac{d\rho}{\rho} = \frac{dP}{K_T} - \alpha dT \quad (10)$$

These equations model a simple convection, with neither phase change nor chemical composition, nor continental crust.

The main consequences of this theory are:

- From observations on the surface (form and geoid), the effects of convection are much weaker than those of rotation.
- The lithospheric plates move, and cold matter sinks within the mantle (observed from seismology) while hot matter rises, producing a weak discrepancy with respect to the hydrostatic equilibrium. The existence of convection requires a temperature distribution different from the adiabatic temperature. On the (geological) time-scale of convection, the mantle has the rheological behavior of a viscous fluid (Newtonian).
- The discrepancy with respect to the hydrostatic equilibrium implies anomalies in the time-variable density; the inertia tensor too varies on geological time-scale.

Theory of the Planets' Interiors, of Deformation, and of Rotation

In terrestrial dynamics, 3 other partial theories deserve mention. The first is the theory of the planets' interiors. It is a static theory which therefore is less complicated than dynamical theories, but it does take into account most of the fundamental thermodynamical equations, potentials, phase changes, equations of state, as well as Poisson's equation. On the other hand, it does not consider the equations of transport and of rheological behavior. We shall not detail this theory, since it does not seem to play any vital role in the study of rotation. The theory has led to a mean model for a radially stratified Earth, the most classical being the PREM derived by Dziewonski and Anderson (1981). The next figure shows density, shear modulus, and modulus of incompressibility as functions of the radius in the PREM.

The remaining 2 partial theories are those of deformation and rotation. These theories will be detailed in the next 2 sections.

Note that these partial theories may be applied, with more or less success, to other telluric planets.

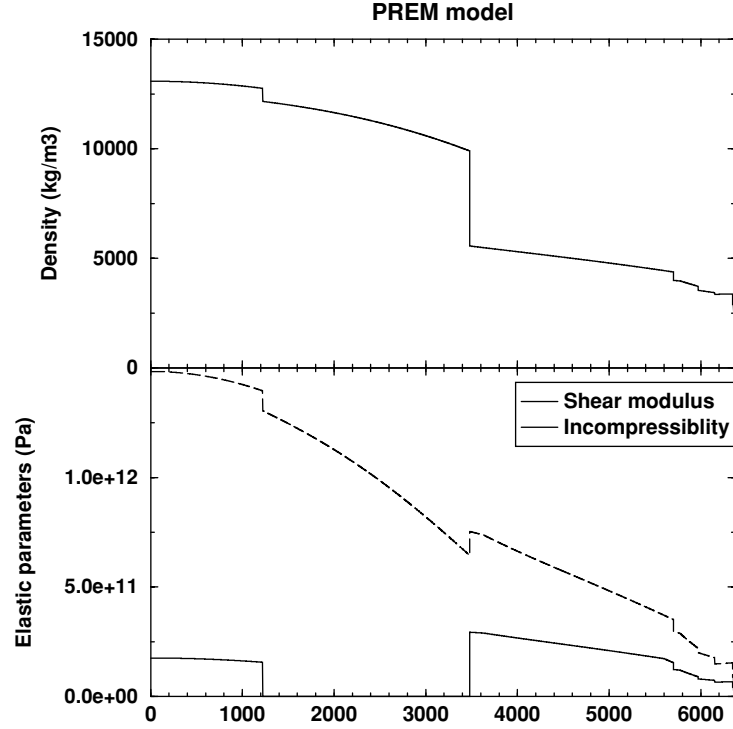


Fig. 1. Top: density (*solid line*). Bottom: shear modulus (*dashed line*) and modulus of incompressibility (*dotted line*) against radius r within the Earth

3 Deformation of a Planet

Come now to the partial dynamical theory of deformation of the Earth or of a planet. Within the general setup of Subsect. 2.2, we keep the conservation of mass and momentum as in the theory of convection, but whereas the equation of convection was for a fluid whose particles could wander far away, in the theory of deformation our equation is for a solid whose particles remain close to their initial positions. Hence, whereas the variable in the equation of convection was the velocity $\mathbf{v}$, the variable in the equation of deformation is the displacement $\mathbf{u}$. For convection the stress tensor was a function of the stress rate. For deformation, it may be written as a function of the stress tensor (elastic case), or as combination of different terms (viscoelastic case). Other important differences from the theory of convection are: thermal effects are weak and so are neglected; the gravitational potential plays a prominent role.

What are the observational data on which the theory of deformation is built?

- seismic elastic modes, on a time-scale less than one hour; we shall not investigate them here
- luni-solar tides on a time-scale ranging from a few hours to a few years
- motion of the rotational pole and evolution of the axial rotation of the Earth
- effects of the time-variable atmospheric or oceanic loadings
- effects of the post-glacial rebound induced by the last deglaciation

These observations show that on a time-scale of seconds to a few years, the Earth has an elastic behavior, whereas on a time-scale over a few hundred years, its rheological behavior is viscoelastic.

After a historical review, we investigate the theory of elastic deformation and then extend the results to viscoelastic deformation in order to study the influence of rheology on the deformation.

3.1 Historical Review

From the viewpoint of this paper, the first questions on the deformation of the Earth were raised in the latter half of the 19th century [16]. Though as early as 1847 Hopkins interpreted seismic vibrations as elastic waves, the chief concern of the problem of deformation was to understand the effects of the luni-solar tidal forces. In 1862 Lord Kelvin made the first calculation on the elastic deformation of a homogeneous incompressible Earth subject to a tidal gravitational potential. This theory predicted that the height of the static oceanic tide (vertical distance between the surface and the seabed) must be different for a rigid Earth and for an elastic Earth. The observation by Darwin followed in 1883: he found $\mu \simeq 80$ GPa. While the hypothesis for a viscous Earth was first proposed in 1865 by Jameson in order to interpret the postglacial rebound in Scandinavia, it was also Darwin who in 1876–1878 carried out theoretical studies on the deformation of a viscous planet. The next breakthrough in the theory of an elastic Earth was due to Newcomb: in 1892, he interpreted the Chandler motion of the rotational pole, observed a year earlier by Chandler, as a perturbation by elasticity of the Eulerian motion of a rigid body.

Of the studies on the deformation of the Earth in the first 30 years of the 20th century, we just mention [16]:

- tidal observations:
 - 1892, Rebeur-Paschwitz with a horizontal pendulum
 - 1913–1914, Schweydar with a gravimeter
- theoretical studies
 - 1905, Herglotz on the elasticity for a 2-layered model of the Earth
 - 1909–1911, Love on the homogeneous compressible model, introducing the Love numbers
 - 1929, Rosenhead on elastic deformations caused by oceans

For the first estimate of the viscosity of the Earth from data on the post-glacial rebound, we have to wait until Haskell, 1935 ($\eta = 10^{21}$ Pa.s). In 1950 Takeuchi numerically obtained a first estimate of the Love numbers by using the seismologists' reference model of the Earth.

Since 1960 the literature on this topic has become immense. Here is a selection of a few major contributions:

- 1959, the system of differential equations by Alterman, Jarosh, Pekeris, the so-called y_i system, which is always used in the present studies
- 1972, Farrell's analysis the deformation of the Earth induced by surface loading
- 1974, two seminal papers: one by Peltier who introduced the Maxwell model of rheology to study the viscoelastic deformations induced by the Pleistocene deglaciation, and another by Smith who took into account spherical asymmetry in the equations
- 1978, Zschau on the influence of dissipation on the Love numbers
- 1980, a formalism by Sasao, Okubo, and Saito relating the deformations of the Earth to the rigid rotation of its fluid core, in order to study the nutation
- starting in the 1980s, a series of papers by Wahr on normal modes in the deformation of the Earth deformation and their application to the model of the Earth as a rotating ellipsoid.

3.2 Elasto-Gravitational Deformation of a Planet

Equations

See for example [12]. We consider a planet made of an elastic mantle and an inviscid fluid core. In the undeformed state, called reference state, the planet is spherical and is described by density ρ_0 , stress σ_0 which is assumed isotropic ($\sigma_0 = -P_0$), and a gravitational potential Φ_0 . We neglect temperature and entropy. This reference state is static. The conservation of momentum says

$$\nabla \cdot \sigma_0 + \rho_0 \nabla \Phi_0 = \mathbf{0} \quad (11)$$

while Poisson's equation says

$$\Delta \Phi_0 = -4\pi G \rho_0 \quad (12)$$

We take solutions of these equations as known, from theory or from seismological observations.

The planet is subject to lunisolar gravity $\rho \mathbf{f}$ with $\mathbf{f} = \nabla V$, where V is the luni-solar tidal potential, or to surface conditions (surface pressure P^e or tangential traction $\mathbf{T}^e$), or to internal surface conditions (pressure P^c or tangential traction $\mathbf{T}^c$). The planet is thereby deformed. The state inside the planet is described by density ρ , displacement $\mathbf{u}$ from the reference state,

strain ($\epsilon_{ij} = \frac{1}{2}(\partial_i u_j + \partial_j u_i)$), stress σ_{ij} , potential due to the rest of the body, and the tidal potential V .

The conservations of mass and momentum, and Poisson's equation, now say

$$\begin{aligned}\partial\rho/\partial t + \nabla \cdot \rho \mathbf{v} &= 0 \\ \rho d^2 \mathbf{u}/dt^2 &= \nabla \cdot \sigma + \rho \nabla \Phi + \rho \nabla V \\ \Delta \Phi &= -4\pi G \rho\end{aligned}\tag{13}$$

Our model is defined by these equations, plus a rheological law relating the stress and strain tensors, which we shall discuss below.

The fundamental assumption in the elasto-gravitational theory is that the deformations are small in comparison with the reference configuration in hydrostatic equilibrium. It is thus natural to use a perturbations theory. Write

$$\begin{aligned}\rho &= \rho_0 + \rho_1^e \\ \sigma &= \sigma_0 + \sigma_1^e \\ \Phi &= \Phi_0 + \Phi_1^e\end{aligned}\tag{14}$$

The velocity $\mathbf{v}$ and the potential V are perturbations. We keep only terms linear in $\mathbf{u}$. Taking (11), (12), we have

$$\begin{aligned}\rho_1^e + \nabla \cdot (\rho_0 \mathbf{u}) &= 0 \\ \rho_0 d^2 \mathbf{u}/dt^2 &= \nabla \cdot \sigma_1^e + \rho_0 \nabla \Phi_1^e + \rho_1^e \nabla \Phi_0 + \rho_0 \nabla V \\ \Delta \Phi_1^e &= -4\pi G \rho_1^e\end{aligned}\tag{15}$$

We now add the rheological law relating σ_1^e and ϵ_{ij} . We have to be careful if the reference state is pre-stressed. If the pre-stress is a pressure (no shear), then the cleanest parametrization of the perturbations of the stress tensor turns out to be the Lagrangian parametrization [12]. We thus introduce the tensor σ_1^l and the rheological law of elasticity for an isotropic medium is:

$$\sigma_{1ij}^l = \lambda (\nabla \cdot \mathbf{u}) \delta_{ij} + 2\mu \epsilon_{ij}\tag{16}$$

with the classical relation between the Lagrangian and Eulerian perturbations:

$$\sigma_1^l = \sigma_1^e + \mathbf{u} \cdot \nabla \sigma_0\tag{17}$$

The conservation of momentum may therefore be written:

$$\rho_0 \frac{d^2 \mathbf{u}}{dt^2} = \nabla \cdot (\lambda (\nabla \cdot \mathbf{u}) + 2\mu \epsilon) + \nabla (\mathbf{u} \cdot \nabla P_0) + \rho_0 \nabla \Phi_1^e + \rho_1^e \nabla \Phi_0 + \rho_0 \nabla V\tag{18}$$

λ and μ are those of the reference model.

Solution in Spherical Representation

The spherical symmetry of the reference model makes it natural to cast the calculations in spherical coordinates and to expand the parameters in spherical harmonics. Let radius r , colatitude θ , longitude φ be the spherical coordinates around the center of mass of the reference model. The spherical harmonics $Y_{nm}^{c,s}(\theta, \varphi)$ satisfy (Heiskanen and Moritz 1967)

$$\Delta \left(r^n Y_{nm}^{c,s}(\theta, \varphi) \right) = 0 \quad (19)$$

where Δ denotes the Laplacian. The spherical harmonics form a basis in which an arbitrary function $f(\theta, \varphi)$ can be expanded:

$$f(\theta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=0}^n \left(a_{nm} Y_{nm}^c + b_{nm} Y_{nm}^s \right) \quad (20)$$

The coefficients a_{nm} and b_{nm} are the spectra of f . A vector version of the spherical harmonics is equally easy: $\nabla Y_{nm}^{c,s}$ the part of the gradient tangent to the sphere, and $\mathbf{e}_r \wedge \nabla Y_{nm}^{c,s}$. An arbitrary vector-valued function $\mathbf{u}(\theta, \varphi)$ can be expanded

$$\mathbf{u}(r, \theta, \varphi) = \sum_{n=0}^{\infty} \sum_{m=0}^n \left(u_{rnm}^{c,s} Y_{nm}^{c,s} \mathbf{e}_r + u_{snm}^{c,s} r \nabla Y_{nm}^{c,s} + u_{tnm}^{c,s} \mathbf{r} \wedge \nabla Y_{nm}^{c,s} \right) \quad (21)$$

$u_{rnm}^{c,s}$, $u_{snm}^{c,s}$, $u_{tnm}^{c,s}$ are respectively the radial, spheroidal, toroidal parts of $\mathbf{u}$. These coefficients are the spectral components of $\mathbf{u}$. For degree $n = 2$, which is very important in the theory of rotation and tides, we have

$$\begin{aligned} Y_{20} &= \frac{3}{2}(\cos^2 \theta - 1) & Y_{21}^c &= 3 \cos \theta \sin \theta \cos \varphi; & Y_{22}^c &= 3 \sin^2 \theta \cos 2\varphi \\ Y_{21}^s &= 3 \cos \theta \sin \theta \sin \varphi; & Y_{22}^s &= 3 \sin^2 \theta \sin 2\varphi \end{aligned} \quad (22)$$

To write the elasto-gravitational equations, we systematically use expansion in spherical harmonics and the differential equations become equations relating the spectral components. The following notation for each spectral component makes the equations look less ferocious:

– for displacements

$$\begin{aligned} u_{rnm}^{c,s} &= y_{1nm}^{c,s} Y_{nm}^{c,s} \\ \mathbf{u}_{snm}^{c,s} + \mathbf{u}_{tnm}^{c,s} &= y_{3nm}^{c,s} r \nabla Y_{nm}^{c,s} + y_{7nm}^{c,s} \mathbf{r} \wedge \nabla Y_{nm}^{c,s} \end{aligned} \quad (23)$$

Hereafter we will write these simply as

$$\begin{aligned} u_{rn} &= y_{1n} Y_n \\ \mathbf{u}_{sn} + \mathbf{u}_{tn} &= y_{3n} r \nabla Y_n + y_{7n} \mathbf{r} \wedge \nabla Y_n \end{aligned} \quad (24)$$

– for traction $\mathbf{T} = \sigma \cdot \mathbf{r}$

$$\begin{aligned} T_{rn} &= y_{2n} Y_n \\ \mathbf{T}_{sn} + \mathbf{T}_{tn} &= y_{4n} r \nabla Y_n + y_{8n} \mathbf{r} \wedge \nabla Y_n \end{aligned} \quad (25)$$

– for gravitational potential

$$\Phi_{1n}^e + V_n = y_{5n} Y_n \quad (26)$$

– for Lagrangian attraction

$$y_{6n} = \dot{y}_{5n} - 4\pi G \rho_0 y_{1n} \quad (27)$$

Using this notation, the equations may be written as a system with 8 first-order equations (the index n being omitted for simplicity). In the frequency (denoted ω) domain [2],

$$\begin{cases} \dot{y}_1 = -\frac{2(K-\frac{2}{3}\mu)}{(K+\frac{4}{3}\mu)} \frac{y_1}{r} + \frac{1}{(K+\frac{4}{3}\mu)} y_2 + \frac{n(n+1)(K-\frac{2}{3}\mu)}{(K+\frac{4}{3}\mu)} \frac{y_3}{r} \\ \dot{y}_2 = \left[-4\rho g - \omega^2 \rho r + \frac{12\mu K}{(K+\frac{4}{3}\mu)r} \right] \frac{y_1}{r} - \frac{4\mu}{(K+\frac{4}{3}\mu)} \frac{y_2}{r} + n(n+1) \left[\rho g - \frac{6\mu K}{(K+\frac{4}{3}\mu)r} \right] \frac{y_3}{r} \\ \quad + \frac{n(n+1)y_4}{r} - \rho y_6 \\ \dot{y}_3 = -\frac{y_1}{r} + \frac{y_3}{r} + \frac{y_4}{\mu} \\ \dot{y}_4 = \left[\rho g - \frac{2\mu(3K)}{(K+\frac{4}{3}\mu)r} \right] \frac{y_1}{r} + \left\{ -\omega^2 \rho r + \frac{2\mu \left[(K-\frac{2}{3}\mu)(2n^2+2n-1) + 2\mu(n^2+n-1) \right]}{(K+\frac{4}{3}\mu)r} \right\} \frac{y_3}{r} \\ \quad - \frac{(K-\frac{2}{3}\mu)}{(K+\frac{4}{3}\mu)} \frac{y_2}{r} - \frac{3y_4}{r} - \rho \frac{y_5}{r} \\ \dot{y}_5 = 4\pi G \rho y_1 + y_6 \\ \dot{y}_6 = -4\pi G \rho n(n+1) \frac{y_3}{r} + \frac{n(n+1)}{r} \frac{y_5}{r} - \frac{2y_6}{r} \\ \dot{y}_7 = \frac{y_7}{r} + \frac{y_8}{\mu} \\ \dot{y}_8 = \left[-\omega^2 \rho r + \frac{\mu(n^2+n-2)}{r} \right] \frac{y_7}{r} - \frac{3y_8}{r} \end{cases} \quad (28)$$

Density $\rho(r)$, rigidity $\mu(r)$, incompressibility $K(r)$, and gravity $g(r)$ depend on the radial stratification of the reference model. This system describes the elasto-gravitational behavior within the elastic parts of a planet. It also allows the study of seismic modes. In studies of deformations with periods much larger than one hour, we classically assume $\omega = 0$: we then speak of static deformations.

The boundary conditions are as follows:

- displacement and attraction vanish at the center of mass $r = 0$
- across each internal interface y_i are continuous
- at surface of the planet $r = a$, the discontinuity in the attraction is equal to the attraction of the deformation bulge

$$y_6(a) + \frac{n+1}{a} y_5(a) = \frac{2n+1}{a} V_n \quad (29)$$

– in the case $n = 1$, we add a condition on the conservation of the center of mass

The solution of the y_i system give the spectral components of the displacement y_1, y_3, y_7 and of the gravitational potential y_5 . Summation recovers the vector $\mathbf{u}$ and the potential Φ_1 .

If there are sources of excitation on a sphere of radius r_0 , they are functions of θ, φ and can be expanded in spherical harmonics or vectors. In this manner, pressure P^e and tangential traction $\mathbf{T}^e$ may be written:

$$\begin{aligned} P^e &= \sum_n P_n^e Y_n \\ \mathbf{T}^e &= \sum_n \left(T_{sn}^e \mathbf{r} \nabla Y_n + T_{tn}^e \mathbf{r} \wedge \nabla Y_n \right) \end{aligned} \quad (30)$$

On the Love Numbers

In many geophysical problems, solutions on the planet's surface or on any internal interface are of special interest. For example, if we assume excitation sources such as volumic potential (with spectral component V_n), surface pressure (component P_n^e), and pressure acting at the core/mantle boundary (component P_n^c), the solutions on the planet's surface are of the form

$$\begin{aligned} y_{1n}(a) &= h_n \frac{V_n}{g_0} + \bar{h}_n \frac{P_n^e}{\rho g_0} + \bar{h}_n^c \frac{P_n^c}{\rho g_0} + \dots \\ y_{3n}(a) &= l_n \frac{V_n}{g_0} + \bar{l}_n \frac{P_n^e}{\rho g_0} + \bar{l}_n^c \frac{P_n^c}{\rho g_0} + \dots \\ \Phi_{1n}(a) &= k_n V_n + \bar{k}_n \frac{P_n^e}{\rho} + \bar{k}_n^c \frac{P_n^c}{\rho} + \dots \end{aligned} \quad (31)$$

The dimensionless numbers h_n, l_n, k_n are called Love numbers; each excitation source has a set of Love numbers. The coefficients ρ and g_0 are normalization factors: ρ is the averaged density and g_0 the surface gravity. The Love numbers are none other than transfer functions for sources of unit excitation.

The Love numbers appear in many geophysical and geodetical observations: gravity measurements, vertical deviation, perturbation in the direction of celestial bodies... Here we detail only one application to perturbations of the inertia tensor [28].

Note first that the external potential of a quasi-spherical body may be written, for degree 2,

$$\begin{aligned} U_2 &= \frac{G}{r^3} \left[Y_{20} \left(\frac{C_{11} + C_{22}}{2} - C_{33} \right) + Y_{21}^c \left(-C_{13} \right) + Y_{21}^s \left(-C_{23} \right) + Y_{22}^c \left(\frac{C_{22} - C_{11}}{4} \right) \right. \\ &\quad \left. + Y_{22}^s \left(-\frac{C_{12}}{2} \right) \right] \end{aligned} \quad (32)$$

C_{ij} are the perturbations from the spherical tensor. Next, for a deformable spherical Earth, the degree 2 potential comes from the deformation: it is just the degree 2 of the gravitational potential Φ_1^e . This potential involves the Love number k_2 . As an example, we put as an excitation source the degree 2 potential V_2 , written at the Earth's surface:

$$V_2 = V_{20}Y_{20} + V_{21}^c Y_{21}^c + V_{21}^s Y_{21}^s + V_{22}^c Y_{22}^c + V_{22}^s Y_{22}^s \quad (33)$$

and consequently

$$\Phi_1^e = k_2 V_2 = U_2 \quad (34)$$

Identifying the coefficients of (32), (34),

$$\begin{aligned} \frac{G}{a^3} \left(\frac{C_{11} + C_{22}}{2} - C_{33} \right) &= k_2 V_{20} \\ -\frac{G}{a^3} C_{13} &= k_2 V_{21}^c & -\frac{G}{a^3} C_{23} &= k_2 V_{21}^s \\ \frac{G}{a^3} \left(\frac{C_{22} - C_{11}}{4} \right) &= k_2 V_{22}^c & -\frac{G}{a^3} \frac{C_{12}}{2} &= k_2 V_{22}^s \end{aligned} \quad (35)$$

Classically these are called MacCullagh's formulae. These concise formulae apply only to perturbations of the global inertia tensor of a planet. To calculate perturbations of a given region, we must go back to the integral over the deformed volume:

$$C_{ij} = \int \rho_1^e (x_m x_m \delta_{ij} - x_i x_j) dv \quad (36)$$

Having established the foundation of the theory of elastic deformation, often called the elasto-gravitational theory, we will now examine some geophysical applications.

Theoretical Body Tides

The most famous application of the Love numbers is the computation of body tides. Let us take the lunar wave M_2 (which has the largest amplitude on the Earth). It is a semi-diurnal wave. The associated degree 2 exciting tidal potential is

$$V_2 = 3 \sin^2 \theta (V_{22} \cos 2\varphi + \tilde{V}_{22} \sin 2\varphi) \quad (37)$$

where $V_{22} = V_0 \cos(2\pi t/T_G)$, $\tilde{V}_{22} = V_0 \sin(2\pi t/T_G)$, $T_G = 12h \ 27mn \ 19s$, the time between 2 successive high tides, $2T_G$ being the time between two passages of the Moon at the local meridian. If the Moon is in the equatorial plane, then $V_0 = 0.8725 \text{ m}^2/\text{s}^2$. The degree 2 tidal Love numbers is computed for the PREM: $h_2 = 0.604$, $k_2 = 0.298$. The radial displacement on the Earth's surface is then:

$$u_r(a, \theta, \varphi) = h_2 \frac{V_2(\theta, \varphi)}{g_0} = h_2 \ 3 \sin^2 \theta \frac{V_0}{g_0} \cos\left(\frac{2\pi}{T_G} t - 2\varphi\right) \quad (38)$$

that is, in Paris ($\theta = 42^\circ$), a radial displacement of about 7.6 cm between high and low tides.

With the same formalism, we can easily compute the perturbation to gravity induced by this tidal potential. On the deformed surface, the gravity perturbation takes the form

$$\delta g(\theta, \varphi) = \frac{2}{a} V_2(\theta, \varphi) \left(1 - \frac{3}{2} k_2 + h_2 \right) = \delta_2 \frac{2}{a} V_2(\theta, \varphi) \quad (39)$$

where δ_2 denotes the degree 2 gravimetric factor. We find a variation of about 50 microgal ($5 \times 10^{-7} \text{m/s}^2$) between successive high and low tides.

Atmospheric Continental Loading

This part addresses the effects of atmospheric loading on surface deformations, using a model proposed by Gegout (1995). The data are the ECMWF record of the pressure field over 11 years. The atmospheric load that has dominant annual and semi-annual components is separated into a part over the continents and a part over the oceans. The load over the oceans does not deform the Earth, because oceans react as an inverted barometer on this time-scale. In contrast the load over the continent deforms the whole Earth, not only the continental surface but also the bottom of the oceans. So we have to take into account variations in the water depth induced by the deformation of the ocean bottom. Using load Love numbers (combination of pressure and tidal Love numbers), the radial and tangential displacements, in Grasse, are computed and plotted in Fig. 2. Their order of magnitude is millimeters.

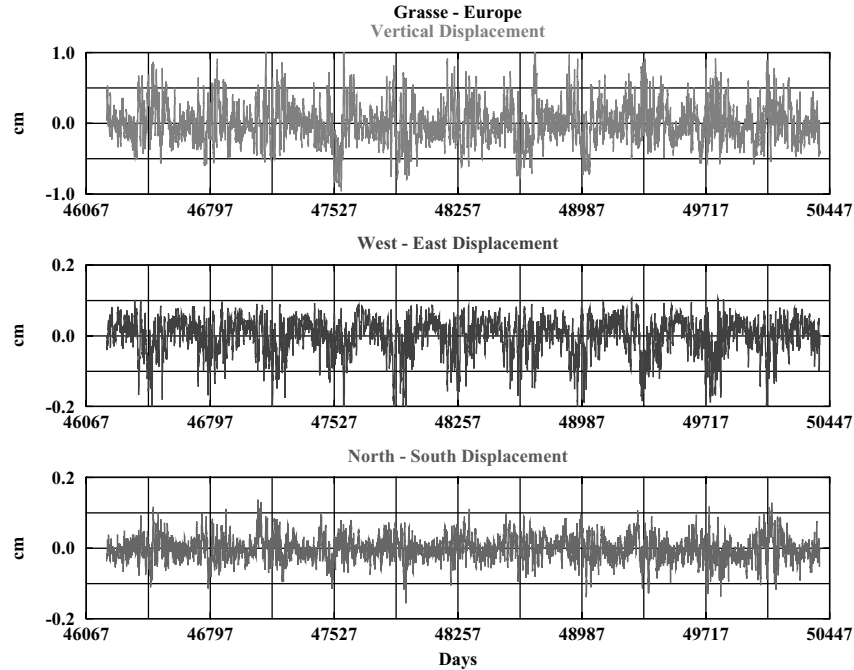


Fig. 2. Displacements in cm, radial (*top*), east-west (*middle*), north-south (*bottom*), computed by P. Gegout (E.O.S.T.)

Elastic Deformations Due to Magnetic Pressure at the Core/Mantle Boundary (CMB)

This magnetic pressure changes at on a scale of 10 years and is estimated from the observations of the surface magnetic field and of its secular variation. Expanding the flow potential in spherical harmonics we first compute, for the last 40 years, the poloidal and toroidal parts of the fluid velocity at the CMB, under the hypothesis of tangential geostrophy. We then compute the associated geostrophic pressure, whose order of magnitude is 1000 Pa. The surface topography induced by this pressure field is computed using Love numbers and is of the order of millimeters. Finally we find the potential for mass redistribution induced by these deformations and, in particular, the zonal components of the surface geopotential (coefficients J_2 , J_3 , J_4): perturbations are about 10^{-10} for J_2 , about 10^{-11} for J_3 , 0.3×10^{-11} for J_4 , varying on a scale of 10 years.

Geostrophic magnetic pressure at the CMB in 1980

from Hulot et al. (1990)

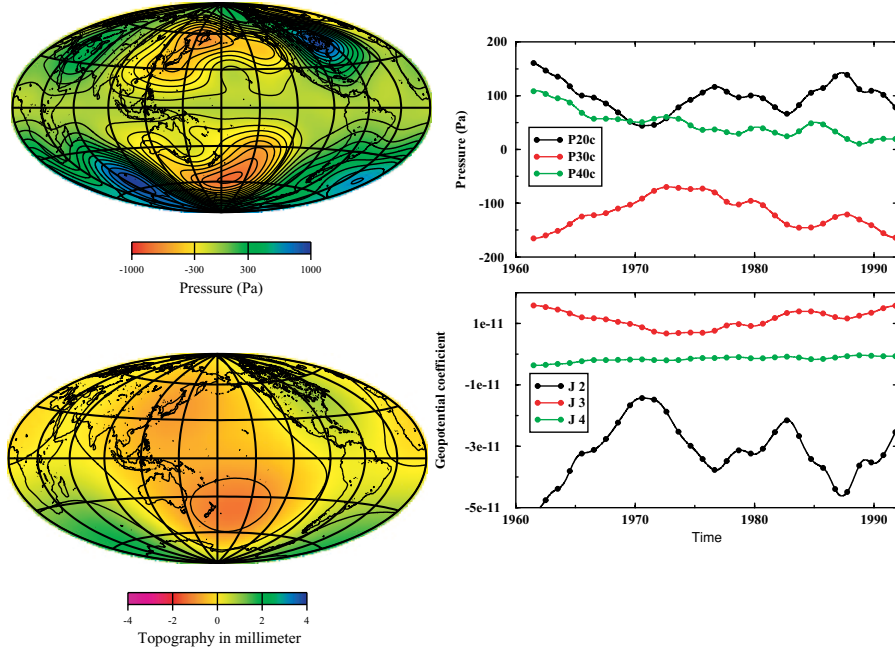


Fig. 3. (a) Excitation function: geostrophic pressure at the CMB in 1980; (b) Surface radial displacement due to elastic deformations induced by the CMB pressure in 1980; (c) Variations in the zonal coefficients of degree 2, 3, 4 of the geostrophic pressure at the CMB; (d) Variations in the associated zonal coefficients of the geopotential

3.3 Viscoelastic Deformation of a Planet

The viscoelastic deformation of the Earth is very similar to the elastic deformation. We have the same physical and mechanical equations (cf 15):

$$\begin{aligned}\rho_1^e + \nabla \cdot (\rho_0 \mathbf{u}) &= 0 \\ \rho_0 \frac{d^2 \mathbf{u}}{dt^2} &= \nabla \cdot \sigma_1^l + \nabla (\mathbf{u} \cdot \nabla P_0) + \rho_0 \nabla \Phi_1^e + \rho_1^e \nabla \Phi_0 + \rho_0 \nabla V \\ \Delta \Phi_1^e &= -4\pi G \rho_1^e\end{aligned}\quad (40)$$

The only change concerns in the rheological law relating stress to strain. Whereas the elastic relation between stress and strain is instantaneous, viscoelasticity has memory: stress depends on the history of strain or vice versa. The parameter of time appears in the rheological law. To express this phenomenon we consider, along with stress and strain tensors, some tensors involving their time-derivatives. Recall that the tensor of strain rate has already been introduced in an earlier section for the convective theory of Newtonian fluids: it was characterized by the parameter of viscosity η .

In this section we explain the elements of rheology and then apply them to the problem of deformations of a planet.

Linear Rheology

We are going to restrict our study to linear rheological behavior [12, 74]. Start with Hooke's law

$$\sigma_{ij} = \lambda(\nabla \cdot \mathbf{u})\delta_{ij} + 2\mu\epsilon_{ij} \quad (41)$$

To characterize a rheological behavior it suffices to consider only one of the Lamé parameters, because the other will have similar properties. We start thus with an even simpler expression

$$\sigma_{ij} = 2\mu\epsilon_{ij} \quad (42)$$

Since time-derivatives of tensors come into play in viscous behaviors, we define linear rheology as a relation of proportionality between tensors and their time-derivatives. For example, with standing for d/dt , we can write:

$$\cdots + a_2 \ddot{\sigma}_{ij} + a_1 \dot{\sigma}_{ij} + a_0 \sigma_{ij} = 2\mu \left(b_0 \epsilon_{ij} + b_1 \dot{\epsilon}_{ij} + b_2 \ddot{\epsilon}_{ij} + \cdots \right) \quad (43)$$

This equation defines the rheological behavior of a particle in a continuous medium. a 's and b 's are constants. This equation is at once simple and practical, because via Fourier transform we can shuttle back and forth between the temporal domain and the frequency domain. The spectral form of the rheological equation is

$$\left(\cdots + (i\omega)^2 a_2 + (i\omega) a_1 + a_0 \right) \sigma_{ij}(\omega) = 2\mu \left(\cdots + (i\omega)^2 b_2 + (i\omega) b_1 + b_0 \right) \epsilon_{ij}(\omega) \quad (44)$$

or

$$\sigma_{ij}(\omega) = 2\mu(\omega)\epsilon_{ij}(\omega) \quad \text{where} \quad \mu(\omega) = \mu \frac{\cdots + (i\omega)^2 b_2 + (i\omega)b_1 + b_0}{\cdots + (i\omega)^2 a_2 + (i\omega)a_1 + a_0} \quad (45)$$

This is Hooke's law, except that stress, strain, and elastic modulus are now functions of the frequency ω . Up to this proviso, the deformation of a viscoelastic planet is described exactly like that of an elastic Earth.

There is much to be said about linear rheology. Let us first decompose (45) into partial fractions:

$$\mu(\omega) = \mu \left(\cdots + i\omega\alpha_1 + \alpha_2 + \sum_{n \geq 3} \frac{\alpha_n}{i\omega + \frac{1}{\tau_n}} \right) \quad (46)$$

where α_3 is a function of τ_3 , α_4 of τ_4 , etc. The denominator $\cdots + (i\omega)^2 a_2 + (i\omega)a_1 + a_0$ has a positive discriminant for linear viscoelastic rheologies and so has real roots $1/\tau_3, 1/\tau_4, \dots$. Their reciprocals $\tau_3, \tau_4, \dots$ are the relaxation times for $\mu(\omega)$. The continuous form of (46) is naturally

$$\mu(\omega) = \mu \left(\cdots + i\omega\alpha_1 + \alpha_2 + \int_{\tau_m}^{\tau_M} f(\tau) \frac{1}{i\omega + \frac{1}{\tau}} \frac{d\tau}{\tau^2} \right) \quad (47)$$

where the integration is over an interval of variation of τ .

Some particular cases of the above formulae are worthy of attention. First, consider the dependence

$$\mu(\omega) = \mu \quad (\alpha_2 = 1) \quad (48)$$

We obtain an elastic medium that obeys Hooke's law.

Second, consider the dependence

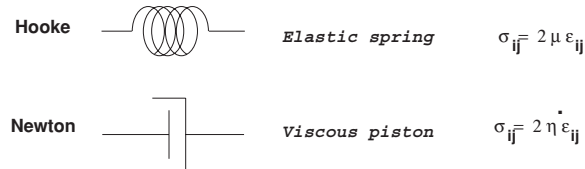
$$\mu(\omega) = \mu\alpha_1 i\omega \quad (49)$$

Noting $\mu\alpha_1 = \eta$, we have

$$\sigma_{ij}(\omega) = 2\eta i\omega \epsilon_{ij}(\omega) \quad \implies \quad \sigma_{ij}(t) = 2\eta \dot{\epsilon}_{ij}(t) = 2\eta d_{ij} \quad (50)$$

We obtain a law of diffusion of velocity. This law defined a Newtonian viscous medium.

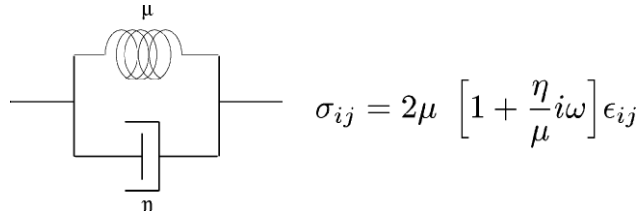
The elastic and viscous behaviors are the two basic rheological behaviors. It is helpful to depict them by a spring and a piston:



We can define more sophisticated rheologies by combining springs and pistons.

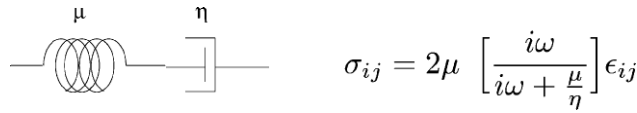
Kelvin Body

If we add stress, we have a connection in parallel, the so-called Kelvin body. It corresponds to the two first terms of equation (46):



Maxwell Body

If we add strain, we have a connection in series, the so-called Maxwell body. It fits into equation (46) with $\alpha_1 = 0$, $\alpha_2 = 1$, $\alpha_3 = -1/\tau_3 = -\mu/\eta$.



The Maxwell body has interesting limits:

- if $\omega \rightarrow 0$, stress disappears and the body reduces to an inviscid fluid;
- if $\omega \rightarrow \infty$, the body becomes elastic;
- if $\omega \ll \mu/\eta$, the body approximates a newtonian viscous fluid.

We can also look at the temporal behavior:

- if $\epsilon_{ij} = H(t)$, then $\sigma_{ij}(t) = 2\mu e^{-\frac{\mu}{\eta}t} H(t)$: relaxation of stress;
- if $\sigma_{ij} = H(t)$, then $\epsilon_{ij}(t) = \frac{\sigma_{ij}}{2\mu} \left(1 + \frac{\mu}{\eta} t \right)$: linear ‘fluage’.

As these illustrations show, the expression for $\mu(\omega)$ in equation (46) gives scope for quite a variety of rheologies. We make one comment on the continuous form (47): setting $\alpha_1 = 0$, $\alpha_2 = 1$, $f(\tau) = \text{const} = -2/\pi Q$ and assuming $\omega\tau_m \ll 1 \ll \omega\tau_M$, we obtain:

$$\mu(\omega) = \mu \left(1 + \frac{2}{\pi Q} \ln(\omega\tau_n) + \frac{i}{Q} \right) \quad (51)$$

a law sometimes used by seismologists.

To end this section, we define two quantities that characterize rheology when the deformation is sinusoidal (caused say by tides): phase delay between the excitation source and the response, and ratio between the dissipated energy and 4π times the averaged energy [54].

Write $\mu(\omega)$ in a complex form $\mu(\omega) = \mu_1 + i\mu_2$. Then for viscoelastic linear rheology, $\mu(-\omega) = \mu_1 - i\mu_2$. Let $\epsilon_{ij}(t) = \epsilon_{ij}^0 \cos \sigma t$. In the frequency domain we have

$$\begin{aligned} \sigma_{ij}(\omega) &= \epsilon_{ij}^0 \left[\mu(-\sigma) \delta(\omega + \sigma) + \mu(\sigma) \delta(\omega - \sigma) \right] \\ &= \epsilon_{ij}^0 \left[\mu_1 \left(\delta(\omega + \sigma) + \delta(\omega - \sigma) \right) + i\mu_2 \left(\delta(\omega - \sigma) - \delta(\omega + \sigma) \right) \right] \end{aligned} \quad (52)$$

In the temporal domain we have:

$$\begin{aligned} \sigma_{ij}(t) &= 2\epsilon_{ij}^0 (\mu_1 \cos \sigma t - \mu_2 \sin \sigma t) = 2\mu_0 \cos(\sigma t + \varphi) \\ \text{with} \quad \mu_1 &= \mu_0 \cos \varphi, \quad \mu_2 = \mu_0 \sin \varphi \end{aligned} \quad (53)$$

The phase delay is then: $\tan \varphi = \mu_2/\mu_1 = \text{Im}\mu(\sigma)/\text{Re}\mu(\sigma)$.

As for the energy, we first compute the dissipative power ΔE over a period T :

$$\begin{aligned} \Delta E &= \int_T \sigma_{ij} \dot{\epsilon}_{ij} dt \text{ with } \dot{\epsilon}_{ij}(t) = \epsilon_{ij}^0 (-\sigma \sin \sigma t) \\ \sigma_{ij}(t) &= \epsilon_{ij}^0 (\mu_1 \cos \sigma t - \mu_2 \sin \sigma t) \end{aligned} \quad (54)$$

Integrating,

$$\Delta E = (\epsilon_{ij}^0)^2 \mu_2 \sigma T = 2\pi \mu_2 (\epsilon_{ij}^0)^2 \quad (55)$$

We compute now the averaged energy $\bar{E}$ over a period T :

$$\bar{E} = \frac{1}{T} \int_T \frac{1}{2} \sigma_{ij}(t) \epsilon_{ij}(t) dt = \frac{1}{2} \mu_1 (\epsilon_{ij}^0)^2 \quad (56)$$

The characteristic parameter is $\Delta E/4\pi\bar{E} = \mu_2/\mu_1$. Note that the two characteristic quantities coincide. Some authors used a factor Q_μ , the reciprocal of the phase delay: $1/Q_\mu = \tan \varphi$.

Now we shall apply linear rheology to the problem of deformations of the Earth.

Love Numbers and Rheology

Mathematically, knowing the Earth's deformation requires integration of the elasto-gravitational equations whose elastic parameters are functions of the frequency and of excitation sources in their turn dependent on the frequency. Deducing the evolution requires convolution of the temporal Love numbers with the temporal sources. The full problem is complicated. We rather choose to bring out the essential structure of the problem by studying a simple example; toward the end we will generalize this example to a more realistic model of a planet [75].

Example: a homogeneous incompressible planet

Consider a planet that is homogeneous, incompressible, and subject to an external gravitational potential. We are interested only in the Love number k_n and we assume the Maxwell model of rheology. It is easy to extend the results we will derive to other sources, other Love numbers, and other rheologies. A homogeneous incompressible model is interesting in that the Love numbers have neat closed forms, and these forms are used in many planetological applications.

For a spectral component of degree n , we have a potential V_n . For the elastic problem, the perturbation of the gravitational potential Φ_{1n}^e on the Earth's surface is given by

$$\Phi_{1n}^e = k_n V_n \quad \text{where } k_n = \text{tidal Love number} \quad (57)$$

Solving the elasto-gravitational equations subject to boundary conditions imposed by an external potential, we find

$$k_n = \frac{3}{2(n-1)} \frac{1}{1 + \frac{2n^2+4n+3}{n} \frac{\mu}{\rho g_0 a}} \quad (58)$$

where μ is the elastic shear modulus, ρ the constant density, g_0 the surface gravity and a the planet radius. In the Maxwell model of rheology,

$$\mu(\omega) = \mu \left(1 - \frac{1}{i\omega\tau + 1} \right) \quad (59)$$

where $\tau = \eta/\mu$ is the Maxwell relaxation time. The factor Q_μ mentioned in the previous section equals $\sigma\tau$. Now go from this local characteristic to the global deformation of the Earth, replacing $\mu(\omega)$ by its expression in (58). We find:

$$k_n(\omega) = k_n^e \left(1 + \frac{\overline{\mu_n}}{\tau_n} \frac{1}{i\omega + \frac{1}{\tau_n}} \right) \quad (60)$$

k_n^e is the elastic Love number, $\overline{\mu_n} = \frac{2n^2+4n+3}{n} \frac{\mu}{\rho g_0 a}$. The degree- n relaxation time $\tau_n = (1 + \overline{\mu_n})\tau$ is different from the Maxwell time. Here we see a difference between local rheology and global deformation. For this simple model, we can easily compute the temporal form of k_n :

$$k_n(t) = k_n^e \left(\delta(t) + \frac{\overline{\mu_n}}{\tau_n} e^{-\frac{t}{\tau_n}} H(t) \right) \quad (61)$$

and we have: $\Phi_{1n}^e(t) = k_n(t) * V_n(t)$, where $*$ denotes the temporal convolution.

We can compute the phase delay of the response in deformation to a sinusoidal excitation source. As in the reasoning we had for rheology, write $V_n(t) = V_n^0 \cos(\sigma t)$ and $k_n(\sigma) = k_{1n} + ik_{2n}$, $k_n(-\sigma) = k_{1n} - ik_{2n}$. The perturbation of the gravitational potential is then

$$\Phi_{1n}^e = V_n^0(k_{1n} \cos(\sigma t) - k_{2n} \sin \sigma t) \quad (62)$$

and the phase delay is $\tan \varphi_{kn} = k_{2n}/k_{1n} = \text{Im}k_n(\sigma)/\text{Re}k_n(\sigma)$.

We can compute the factor $\Delta E/4\pi\bar{E} = 1/Q_E$ too. The elastic energy is

$$E = \frac{1}{2} \left(\int_S T_i u_i ds + \int_V f_i u_i dv \right) \quad (63)$$

When the surface of the Earth is free (no external pressure), the radial traction is equal to the pressure due to the deformed bulge: $T_r = -\rho g_0 u_r$. For an homogeneous body we have

$$\int_V \rho \nabla(\Phi_1^e + V) \cdot \mathbf{u} dv = \int_S \rho(\Phi_1^e + V) u_r ds \quad (64)$$

Thus the energy may be written

$$E = \frac{1}{2} \int_S \left(-\rho g_0 u_r u_r + \rho(\Phi_1^e + V) u_r \right) ds \quad (65)$$

or, using Love numbers, in the frequency domain:

$$E = \frac{\rho}{2g_0} \int_S h_n V_n (1 + k_n - h_n) V_n ds \quad (66)$$

For a viscoelastic body, the Love numbers are time-dependent and so we have to convolve to compute the energy:

$$\begin{aligned} \Delta E &= \frac{\rho}{g_0} \int_T \int_S \left[\left(\delta(t) + k_n(t) - h_n(t) \right) * V_n(t) \right] \left[h_n(t) * \dot{V}_n(t) \right] dt ds \\ \bar{E} &= \frac{\rho}{2g_0 T} \int_T \int_S \left[\left(\delta(t) + k_n(t) - h_n(t) \right) * V_n(t) \right] \left[h_n(t) * V_n(t) \right] dt ds \end{aligned} \quad (67)$$

The Ansatz $V_n(t) = V_n^0 \cos \sigma t$, $k_n(\sigma) = k_{1n} + ik_{2n}$, $h_n(\sigma) = h_{1n} + ih_{2n}$ gives

$$\begin{aligned} \Delta E &= \frac{\pi \rho}{g_0} \left[h_{1n} k_{2n} - h_{2n} (1 + k_{1n}) \right] \int_S (V_n^0)^2 ds \\ \bar{E} &= \frac{\rho}{4g_0} \left[h_{1n} (1 + k_{1n} - h_{1n}) + h_{2n} (k_{2n} - h_{2n}) \right] \int_S (V_n^0)^2 ds \end{aligned} \quad (68)$$

whence

$$\frac{\Delta E}{4\pi\bar{E}} = \frac{h_{1n} k_{2n} - h_{2n} (1 + k_{1n})}{h_{1n} (1 + k_{1n} - h_{1n}) + h_{2n} (k_{2n} - h_{2n})} = \frac{1}{Q_E} \quad (69)$$

The factor Q is different depending on whether we are interested in the phase delay or in the dissipated energy. For the Maxwell model of rheology,

$$\begin{aligned} \tan \varphi_{kn} &= -\frac{\bar{\mu}_n \sigma \tau}{1 + (1 + \bar{\mu}_n) \sigma^2 \tau^2} \\ \frac{\Delta E}{4\pi\bar{E}} &= \frac{\bar{\mu}_n \sigma \tau}{(1 + k_n^e - h_n^e)(1 + (1 + \bar{\mu}_n) \sigma^2 \tau^2) + \bar{\mu}_n (k_n^e - h_n^e)} \end{aligned} \quad (70)$$

where e indicates the elastic Love numbers.

For a homogeneous incompressible Earth we have:

$$k_n^e = \frac{3}{2(n-1)} \frac{1}{1 + \bar{\mu}_n}, \quad h_n^e = \frac{5}{2(n-1)} \frac{1}{1 + \bar{\mu}_n} \quad (71)$$

hence

$$\frac{\Delta E}{4\pi \bar{E}} = \frac{(n-1)\sigma\tau\bar{\mu}_n}{(n-2)(1 + \sigma^2\tau^2) + \bar{\mu}_n(n-1)\sigma^2\tau^2} \quad (72)$$

Note that for $n = 1$, $1/Q_E$ vanishes: degree 1 represents rigid translation and consequently there is no viscoelastic effect and no dissipated energy. Another interesting case is $n = 2$: this degree is very important for the study of the inertia tensor and of the Earth's rotation. For $n = 2$, the equation (72) simplifies and leads to $1/Q_E = 1/\sigma\tau = 1/Q_\mu$.

The formulae (70) for the Maxwell model of rheology deserve a remark or two:

- If we put $\tan \varphi_{kn} = 1/Q_{kn}$, note the factor Q_{kn} becomes negative. To abide by the classical convention on the dissipative factor Q , we will change the definition slightly:

$$\frac{1}{Q_{kn}} = |\tan \varphi_{kn}| \quad (73)$$

Similarly for $1/Q_E$.

- Note that the approximations to (70) will look different depending on whether $\sigma\tau \ll 1$ or $\sigma\tau \gg 1$.

It was pointed out that seismologists use the rheology defined by (51). If we adopt $\mu(\omega)$ of (51) and substitute it in the expression of k_n (58), we get

$$k_n(\omega) = k_n^e \frac{1}{1 + \frac{\bar{\mu}_n}{1 + \bar{\mu}_n} \frac{2}{\pi Q} \ln(\omega\tau_m) + i \frac{\bar{\mu}_n}{1 + \bar{\mu}_n} \frac{1}{Q}} \quad (74)$$

When $Q \gg 1$, the binomial approximation gives the following result, classical in the literature:

$$k_n(\omega) = k_n^e \left(1 - \frac{\bar{\mu}_n}{1 + \bar{\mu}_n} \frac{2}{\pi Q} \ln(\omega\tau_m) - i \frac{\bar{\mu}_n}{1 + \bar{\mu}_n} \frac{1}{Q} \right) \quad (75)$$

In this case $\frac{1}{Q_{kn}} = \frac{\bar{\mu}_n}{1 + \bar{\mu}_n} \frac{1}{Q}$.

A Possible Generalization

Let us upgrade the simple example of a homogeneous incompressible planet with the Maxwell model of rheology to a more realistic model of a stratified compressible planet with a more complex rheology. Peltier (1974) showed that the Love number k_n (and the other Love numbers) may be written in a generic form

$$k_n(\omega) = k_n^e \left(1 + \sum_p \sum_j \frac{r_{pj}}{i\omega + \frac{1}{\tau_{pj}}} \right) \quad (76)$$

where $\sum_p$ runs over relaxation modes related to the nature of the rheology, while $\sum_j$ runs over relaxation modes related to the stratification of the planet, the difference between two layers in the model being characterized by discontinuities in density, viscosity, or elastic modulus.

Geophysical Application: Last Pleistocenic Deglaciation and Post-Glacial Rebound

A classical application of the viscoelasto-gravitational theory is to deformations caused by the last glaciation/deglaciation cycle. The glaciation/deglaciation is modeled by three spherical sheets of ice which are analysed in spherical harmonics with the same time-dependence for the evolution of their respective heights. Glaciation builds up over a long period of $T_0 = 9 \times 10^4$ years but deglaciation is precipitated in a short period of $T_1 - T_0 = 10^4$ years (Fig. 4-a).

The variation in the sea level concomitant with the variation in the height of the ice sheets is taken into account and translates into a surface loading, which is analysed in spherical harmonics. Figure 4-b plots the anomaly in surface density at the end of glaciation.

Using the viscoelastic Love numbers, we compute the evolution of the CMB topography induced by this surface load: 5×10^4 years after the beginning of glaciation (Fig. 4-c), at the end of glaciation (Fig. 4-d), at the end of deglaciation (Fig. 4-e), and at present (i.e. 6000 years after the end of deglaciation) when the topography is of the order of ten meters (Fig. 4-f). Figure 5-a plots the present topography of the ICB: it is a topography with a large wavelength of the order of meters. The present rate of radial displacement (vertical motion) on the Earth's surface induced by the last deglaciation is plotted in Fig. 5-b: it is about a few millimeters per year.

These topographies were computed assuming a viscosity contrast of 30 between the lower and upper mantle. To illustrate to what extent the computed results depend on this assumption, we have computed the present time-derivative of the zonal degree 2 coefficient $\dot{J}_2$ of the geopotential as a function of the lower mantle viscosity (Fig. 6).

The last deglaciation explains the observed value of $\dot{J}_2$ in two models: a model with a quasi-isoviscous mantle and a model with a large viscosity contrast between the lower and upper mantle. The latter model is in good agreement with the geodynamical studies on how heterogeneities in the mantle density influences the geoid.

3.4 Perspectives

Let us review the recent progress on the theoretical foundation of elastic and viscoelastic deformations of the Earth.

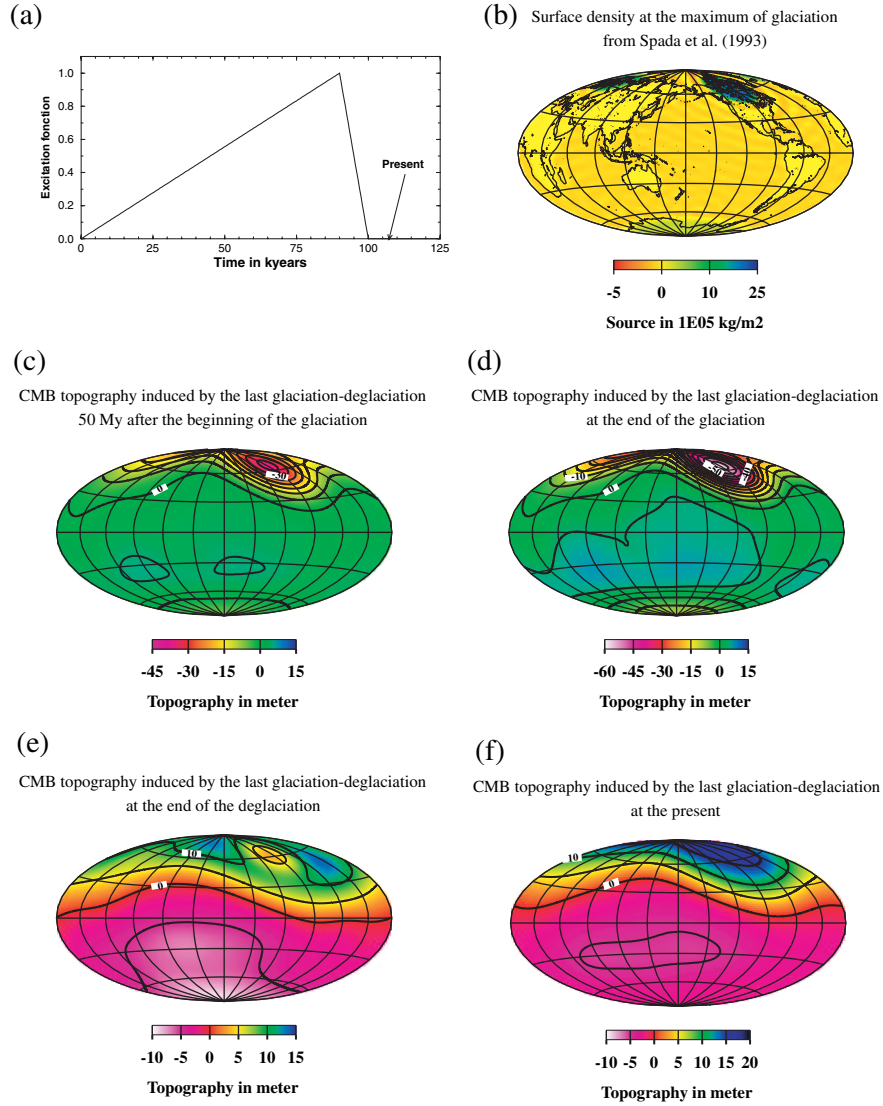
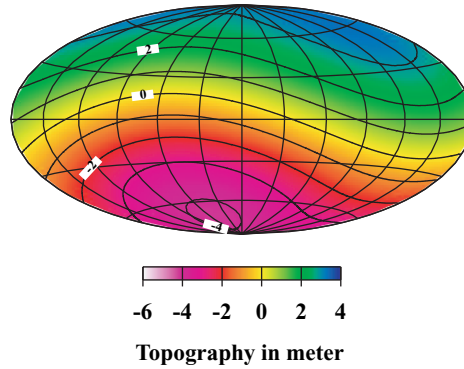


Fig. 4. (a) Evolution of glaciation-deglaciation cycle. (b) Surface density in kg/m^2 at the end of glaciation. (c) CMB topography 5×10^4 years after the beginning of glaciation. (d) CMB topography at the end of glaciation. (e) CMB topography at the end of deglaciation. (f) Present-day CMB topography

For elastic deformations we started from a spherical reference model in hydrostatic equilibrium. But of course the Earth is an ellipsoid and it is not exactly in hydrostatic equilibrium, because of convection within the mantle. The problem then is, how do the ellipsoidal form, tectonic stresses,

(a) ICB topography induced by the last glaciation-deglaciation at the present



(b) Present radial displacement rate

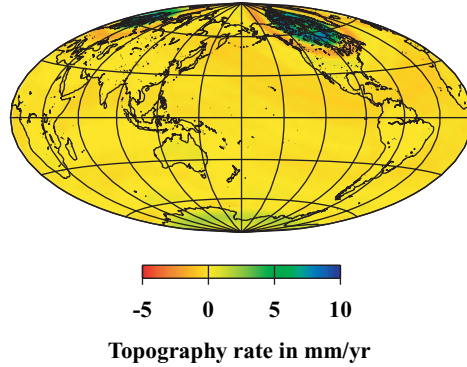


Fig. 5. (a) Present ICB topography induced by the last deglaciation. (b) Present-day rate of radial deformation on the Earth's surface: maxima are under the ice-sheets (prominently under Scandinavia)

lateral anomalies in density and elastic modulus perturb the deformation of the Earth?

For viscoelastic deformations, the major current question concerns how to model rheology, most acutely for phenomena on a time-scale of hours to years. Is it possible to construct a model of rheology between the seismological rheology and the Maxwell rheology? At present the constraints offered by the observations (about nutations, tides, Chandler motion...) are too weak and not conclusive enough (because of the influence of the oceans which is poorly understood) to answer to the question.

One last point concerns relaxation modes. They are easily determined when the planet is radially stratified in homogeneous layers, but their

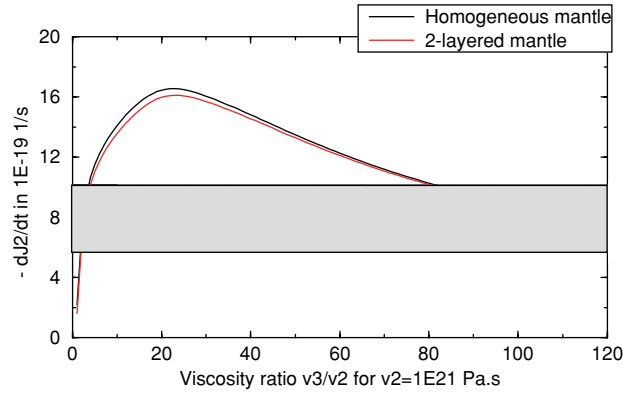


Fig. 6. $\dot{J}_2$ for lower mantle viscosity of 10^{21} to 10^{23} Pa.s. The *shaded area* is the observed value and its error bar (Cheng et al. 1989)

determination is still beyond reach in case of continuous variations of the physical parameters within the planet.

So, work must go on...

4 Rotation of a Deformable Stratified Planet

The theory of rotation of a rigid planet is classical. The principal results are

- the Earth has an axial rotation
- the rotational axis itself moves within the planet: this is the Eulerian motion
- the axis moves with respect to the celestial reference frame: these are precession, nutation, ...

We will review, in this last part, changes brought about in the theory of rotation of a rigid planet by the consideration of a more realistic model of the Earth, in which the Earth is able to be elastically or viscoelastically deformed and is stratified in a solid mantle, a fluid core, and a solid inner core. We will detail the problems for elastic deformations, merely sketching the results on the secular deceleration of a viscoelastic Earth and the large deviation of the rotational pole on a geological time-scale.

4.1 Historical Review

After remarks by Newton (1687), the theory of rotation of a rigid body was initiated in 1749 by d'Alembert and Euler. Euler even proposed a denser core to account for the value of the precessional constant, though a full study of the influence of the internal properties of the Earth on its rotation came only

later. In 1758 Euler went on to discover the Eulerian motion of the pole of rotation.

In 1839 Hopkins investigated the influence that a fluid part within the Earth would have on the value of the precessional constant. This date may be regarded as the starting point of the studies on this topic. For the effects of viscosity or of anelasticity, Darwin (1876–1878) Schiaparelli did pioneering work; the problem was to understand possible large deviations of the rotation axis. Large rotational shift was a recurrent hypothesis by geologists to explain climatic variations and sometimes tectonic phenomena. In 1895, Hough computed more precisely the influence of a fluid core on the Earth's rotation, and in 1910 Poincaré, in his paper on the precession of deformable bodies, laid the foundation of the present theory. But this study fell into oblivion until 1957 when Jeffreys and Vicente computed the influence of a fluid core on the rotation of an elastic deformable Earth. Improved by Molodensky in 1961, this theory took its classical form in the work by Sasao, Okubo, and Saito in 1980.

There were isolated works by Milankovitch in 1934 and Jeffreys in 1952, but the problem on large deviations owes its revival to Gold in 1955 and Inglis in 1957. Their line of research led to present work on the secular shift of the rotational pole induced by the last deglaciation, and the polar wander due to the evolution of anomalies in mantle density (this may be associated with the observed True Polar Wander of the geomagnetic dipolar axis).

The problem of secular deceleration, first considered by Kant and Darwin, is fraught with obscurities. The theoretical problem of viscous rheology of the Earth [36] is still unsolved, and the influence of the oceans, a problem raised by Thomson and Tait in 1879, remains unclear. However, observations of this deceleration have improved dramatically of late: data come from receding of the Moon and from geological observations. For details, cf. [39, 51].

For further historical information cf. also [16].

4.2 Rotation with a Fluid Core and a Solid Inner Core

Seismological Remark

Seismologists investigate elastic vibrations of the Earth. They work on the wave propagation from a source (P and S wave velocities), as well as on the seismic eigenmodes for a closed Earth. We would like to insist on this latter point. The eigenmodes are oscillating modes of the displacement field governed by the elasto-gravitational equations. There are spheroidal and toroidal modes corresponding to the expansion in spherical harmonics. We are interested in the toroidal modes $\mathbf{u}_n$ (spectral component of degree n):

$$\mathbf{u}_n = y_{7n}(r) \mathbf{r} \wedge \nabla Y_n(\theta, \varphi) \quad (77)$$

These are torsional displacements in a tangential plane at each sphere of radius r . In particular for degree 1,

$$Y_1(\theta, \varphi) = \omega_3 \cos \theta + \omega_1 \sin \theta \cos \varphi + \omega_2 \sin \theta \sin \varphi \quad (78)$$

with constants $\omega_1, \omega_2, \omega_3$. Since $rY_1(\theta, \varphi) = \omega_3 z + \omega_1 x + \omega_2 y$, the displacement field may be written $\mathbf{u} = \boldsymbol{\omega} \wedge \mathbf{r}$. This is a rigid rotation.

Seismological calculations showed that there exist three distinct modes of rotation: one mode $\boldsymbol{\omega}$ associated with the Earth as a whole, one mode $\boldsymbol{\omega}^c$ associated with the differential rotation between the fluid core and the Earth as a whole, and one mode $\boldsymbol{\omega}^{ic}$ associated with the differential rotation between the inner core and the Earth as a whole. These rotational modes represent the ‘eigendegrees of freedom’, and historically it was from this point of view that the rotation of the fluid core was investigated. Mechanically this makes sense: if we fill a spherical shell with fluid and spin the shell, then the fluid, retarded by inertia, will have a relative motion with respect to the shell. Similar reasoning applies to a solid floating in a spherical fluid. Thus we can analyse the rotational modes by the seismic method or by the mechanical method (Liouville equation). Either way, the key is the coupling between the fluid core and the solid mantle or the solid inner core. Hereafter we will use the mechanical method.

Equation of Rotation

The equations governing the angular velocities $\boldsymbol{\omega}, \boldsymbol{\omega}^c, \boldsymbol{\omega}^{ic}$ are derived from the angular momentum theorem. Let I^m, I^c, I^{ic} be the inertia tensors of the mantle, of the fluid core, of the solid inner core. Then their angular momenta are $\mathbf{H}^m = I^m \boldsymbol{\omega}, \mathbf{H}^c = I^c(\boldsymbol{\omega} + \boldsymbol{\omega}^c), \mathbf{H}^{ic} = I^{ic}(\boldsymbol{\omega} + \boldsymbol{\omega}^{ic})$, and the total angular momentum $\mathbf{H}^t$ of the whole Earth is

$$\mathbf{H}^t = \mathbf{H}^m + \mathbf{H}^c + \mathbf{H}^{ic} \quad \text{or} \quad H_{ij}^t = I_{ij}^m \omega_j + I_{ij}^c \omega_j^c + I_{ij}^{ic} \omega_j^{ic} \quad (79)$$

where $I_{ij} = I_{ij}^m + I_{ij}^c + I_{ij}^{ic}$ is the inertia tensor of the whole Earth. Now apply the angular momentum theorem to $\mathbf{H}^t, \mathbf{H}^c, \mathbf{H}^{ic}$. In the frame attached to the mantle and rotating at angular velocity $\boldsymbol{\omega}$,

$$\begin{cases} \frac{d\mathbf{H}^t}{dt} + \boldsymbol{\omega} \wedge \mathbf{H}^t = \boldsymbol{\Gamma}^t \\ \frac{d\mathbf{H}^c}{dt} + \boldsymbol{\omega} \wedge \mathbf{H}^c = \boldsymbol{\Gamma}^c \\ \frac{d\mathbf{H}^{ic}}{dt} + \boldsymbol{\omega} \wedge \mathbf{H}^{ic} = \boldsymbol{\Gamma}^{ic} \end{cases} \quad (80)$$

the right sides of being the torques acting on the whole Earth, on the fluid core, and on the solid inner core.

Linearized Equation of Rotation

These equations can be simplified by some assumptions justified for the terrestrial rotation.

- The Earth is axisymmetric in its initial reference state. In this state the inertia tensors have the form

$$I_0 = \begin{pmatrix} A & 0 & 0 \\ 0 & A & 0 \\ 0 & 0 & C \end{pmatrix}, \quad I_0^c = \begin{pmatrix} A^c & 0 & 0 \\ 0 & A^c & 0 \\ 0 & 0 & C^c \end{pmatrix}, \quad I_0^{ic} = \begin{pmatrix} A^{ic} & 0 & 0 \\ 0 & A^{ic} & 0 \\ 0 & 0 & C^{ic} \end{pmatrix} \quad (81)$$

- In the initial reference state all the components of the angular velocities vanish except the axial component Ω of $\boldsymbol{\omega}$.
- The rotation of the Earth is a perturbation of its initial reference state and the forces and the torques are of the order of this perturbation. Consequently, denoting the components of the angular velocities

$$\boldsymbol{\omega} = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \Omega + \omega_3 \end{pmatrix}, \quad \boldsymbol{\omega}^c = \begin{pmatrix} \omega_1^c \\ \omega_2^c \\ \omega_3^c \end{pmatrix}, \quad \boldsymbol{\omega}^{ic} = \begin{pmatrix} \omega_1^{ic} \\ \omega_2^{ic} \\ \omega_3^{ic} \end{pmatrix} \quad (82)$$

we assume $\omega_i \ll \Omega$, $\omega_i^c \ll \Omega$, $\omega_i^{ic} \ll \Omega$ for $i = 1, 2, 3$. We also assume that perturbations C_{ij} on the inertia tensors are small compared with the initial principal moments of inertia.

The perturbations on the inertia tensors come, on one hand, from the deformation caused by the rotation itself, and on the other hand from the deformations due to the acting forces (cf. section on the theory of deformations). The torques come from body forces, tidal or gravitational, or from surface forces.

This is a linear perturbation theory: products of perturbation are neglected.

Comments on the Fluid Core

The application of the angular momentum theorem to the solid part is classical. Its application to the fluid part presents some points of interest. If we consider an inviscid fluid core,

$$\rho_0 \boldsymbol{\gamma} = -\nabla P_1^e + \rho_0 \nabla (\Phi_1^e + V) \quad (83)$$

where P_1^e is the Eulerian perturbation of the pressure taking into account the effects of rotation. The acceleration $\boldsymbol{\gamma}$ is, in linear approximation,

$$\boldsymbol{\gamma} = \frac{d}{dt}(\boldsymbol{\omega} + \boldsymbol{\omega}^c) \wedge \mathbf{r} + 2(\boldsymbol{\omega} + \boldsymbol{\omega}^c) \wedge (\boldsymbol{\omega}^c \wedge \mathbf{r}) + \boldsymbol{\omega} \wedge (\boldsymbol{\omega} \wedge \mathbf{r}) + \mathbf{r}(\boldsymbol{\omega}^c \cdot \boldsymbol{\omega}^c) - \boldsymbol{\omega}^c(\boldsymbol{\omega}^c \cdot \mathbf{r}) \quad (84)$$

Assume the fluid core to be barotropic. From the equilibrium equation we have $\nabla \wedge \boldsymbol{\gamma} = \mathbf{0}$ and the relation $\frac{d}{dt}(\boldsymbol{\omega} + \boldsymbol{\omega}^c) = \boldsymbol{\omega}^c \wedge \boldsymbol{\omega}$, which allows us to rewrite the acceleration as

$$\boldsymbol{\gamma} = \boldsymbol{\omega} \left[(\boldsymbol{\omega} + \boldsymbol{\omega}^c) \cdot \mathbf{r} \right] + \boldsymbol{\omega}^c \left[(\boldsymbol{\omega} + \boldsymbol{\omega}^c) \cdot \mathbf{r} \right] - \mathbf{r} \left[(\boldsymbol{\omega} + \boldsymbol{\omega}^c) \cdot (\boldsymbol{\omega} + \boldsymbol{\omega}^c) \right] \quad (85)$$

The structure of the acceleration reveals:

- γ is a gradient in linear approximation:

$$\gamma = \nabla(\Pi + \Psi) \quad \text{with} \quad \begin{aligned} \Pi &= -\frac{\Omega\omega_1^c}{3}r^2Y_{21}^c - \frac{\Omega\omega_2^c}{3}r^2Y_{21}^s \\ \Psi &= -\frac{\Omega^2}{3}r^2Y_{20} - \frac{\Omega\omega_1}{3}r^2Y_{21}^c - \frac{\Omega\omega_2^c}{3}r^2Y_{21}^s \end{aligned} \quad (86)$$

where Ω is the axial sidereal rotation.

- The angular momentum theorem yields

$$\frac{d\mathbf{H}^c}{dt} + \boldsymbol{\omega} \wedge \mathbf{H}^c = \int_{\text{core}} \mathbf{r} \wedge (-\nabla P_1^e + \rho \nabla(\Phi_1^e + V)) dv = \int_{\text{core}} \mathbf{r} \wedge \rho \gamma dv \quad (87)$$

From the expression of γ in linear approximation (85)

$$\int_{\text{core}} \mathbf{r} \wedge \rho \gamma dv = (\boldsymbol{\omega} + \boldsymbol{\omega}^c) \wedge \mathbf{H}^c \quad (88)$$

and therefore

$$\frac{d\mathbf{H}^c}{dt} + \boldsymbol{\omega} \wedge \mathbf{H}^c = (\boldsymbol{\omega} + \boldsymbol{\omega}^c) \wedge \mathbf{H}^c \quad \text{or} \quad \frac{d\mathbf{H}^c}{dt} - \boldsymbol{\omega}^c \wedge \mathbf{H}^c = \mathbf{0} \quad (89)$$

We recover the expression for the Helmholtz vortex discovered by Poincaré.

- The previous expression shows that the torque due to external body forces vanishes on the fluid core because these forces are canceled by pressure effects existing at the boundary.
- The fluid motion within the core is affected by boundary layers along the core/mantle boundary (CMB) and the inner core boundary (ICB), which account for viscous or electromagnetic couplings between the fluid core and the solid parts. A torque appears, perturbing the angular momentum:

$$\frac{d\mathbf{H}^c}{dt} - \boldsymbol{\omega}^c \wedge \mathbf{H}^c = \mathbf{T}^c = \int_S \mathbf{r} \wedge \mathbf{F}(\omega_1^c, \omega_2^c) ds \quad (90)$$

where $\mathbf{F}$ is the viscomagnetic force at the boundaries of the fluid core. The magnetic couplings are due to interactions between the magnetic field and the electric currents induced within the conducting mantle and the inner core by the differential rotation of the fluid core. There is in addition a viscomagnetic torque acting on the inner core, denoted $\mathbf{T}^{ic}$.

Complex Form of Equations

We can compute $\boldsymbol{\omega}$, $\boldsymbol{\omega}^c$, $\boldsymbol{\omega}^{ic}$ in linear approximation from the two angular momentum equations for the whole Earth and the inner core, plus the Helmholtz equation for the fluid core.

Note that the axial equations are decoupled from the equatorial equations. For the axial rotation, we obtain 3 secular modes of the whole Earth, of the fluid core, and of the inner core, in the absence of external or surface

torques. The equatorial equations are best handled in a complex form. For the equatorial components of the angular velocities, for perturbations on the inertia tensor, and for the acting torques,

$$\begin{aligned}
 \omega &= \omega_1 + i\omega_2 & \mathbf{C} &= C_{13} + iC_{23} & \Gamma^t &= \Gamma_1^t + i\Gamma_2^t \\
 \omega^c &= \omega_1^c + i\omega_2^c & \mathbf{C}^c &= C_{13}^c + iC_{23}^c & T^c &= T_1^c + iT_2^c \\
 \omega^{ic} &= \omega_1^{ic} + i\omega_2^{ic} & \mathbf{C}^{ic} &= C_{13}^{ic} + iC_{23}^{ic} & T^{ic} &= T_1^{ic} + iT_2^{ic} \\
 & & & & \Gamma^{ic} &= \Gamma_1^{ic} + i\Gamma_2^{ic}
 \end{aligned} \tag{91}$$

The equatorial part of the angular momentum equations (80) take the form

$$\begin{cases}
 A\dot{\omega} - i\alpha A\Omega\omega + \dot{\mathbf{C}}\Omega + i\Omega^2\mathbf{C} + A^c\dot{\omega}^c + A^{ic}\dot{\omega}^{ic} + i\Omega(A^c\omega^c + A^{ic}\omega^{ic}) = \Gamma^t \\
 A^c\dot{\omega} + A^c\dot{\omega}^c + i(1 + \alpha^c)A^c\Omega\omega^c + \dot{\mathbf{C}}^c\Omega = T^c \\
 A^{ic}\dot{\omega} + A^{ic}\dot{\omega}^{ic} - i\alpha^{ic}\Omega A^{ic}\omega + \dot{\mathbf{C}}^{ic}\Omega + i\Omega^2\mathbf{C}^{ic} + i\Omega A^{ic}\omega^{ic} = \Gamma^{ic} + T^{ic}
 \end{cases} \tag{92}$$

α , α^c , α^{ic} are respectively the dynamic flattening of the Earth ($\alpha = \frac{C-A}{A}$), of the fluid core, and of the inner core. α , α^c , $\alpha^{ic} \ll 1$. For expressions in complex form of perturbations on the products of inertia and of various torques, cf. Greff, Legros, and Dehant (2000).

Deformations and Inertia Tensor Perturbations

The Earth, the fluid core, and the inner core are deformed not only by the tidal potential, but also by the rotation itself and by the load associated with tilting of the inner core. We write

- Tidal potential: $V = (V_{21}Y_{21}^c + \tilde{V}_{21}Y_{21}^s)\left(\frac{r}{a}\right)^2$ for $r \leq a$
- Rotational potential of the Earth: $V^r = \left(-\frac{\Omega\omega_1 a^2}{3}Y_{21}^c - \frac{\Omega\omega_2 a^2}{3}Y_{21}^s\right)\left(\frac{r}{a}\right)^2$ for $r \leq a$
- Rotational potential of the fluid core: $V^c = \Pi = \left(-\frac{\Omega\omega_1^c b^2}{3}Y_{21}^c - \frac{\Omega\omega_2^c b^2}{3}Y_{21}^s\right)\left(\frac{r}{b}\right)^2$ for $c \leq r \leq b$
- Rotational potential of the inner core: $V^{ic} = \left(-\frac{\Omega\omega_1^{ic} c^2}{3}Y_{21}^c - \frac{\Omega\omega_2^{ic} c^2}{3}Y_{21}^s\right)\left(\frac{r}{c}\right)^2$ for $r \leq c$.
- Tilting of the ellipsoidal inner core (expressed by the displacement $u_s = \frac{2}{3}c\alpha^{ic}(-\theta_2^{ic}Y_{21}^c + \theta_1^{ic}Y_{21}^s)$, with $\partial\theta_i^{ic}/\partial t = \omega_i^{ic}$) induces an internal loading potential S^{ic} because of the density contrast across the ICB:

$$S^{ic} = \frac{4\pi G}{5}(\rho^{ic} - \rho^c(c))cu_s \begin{cases} \frac{r^2}{c^2} & \text{if } r \leq c \\ \frac{c^3}{r^3} & \text{if } r \geq c \end{cases} \tag{93}$$

where ρ^{ic} is the density of the inner core and $\rho^c(c)$ that of the outer core at the ICB.

Solving the elasto-gravitational equations [2] subject to the boundary conditions described in the previous section, we compute Love numbers such as:

$$\begin{cases} y_1(c^-, \theta, \varphi) = \frac{c}{a} m \frac{V^r + V}{g_0} + \frac{ac}{b^2} m_1 \frac{V^c}{g_0} + \frac{a}{c} m_2 \frac{V^{ic}}{g_0} + h'_s \frac{S^{ic}}{g_0} + u_s \\ y_5(a, \theta, \varphi) = (1+k)(V^r + V) + \frac{a^2}{b^2} k_1 V^c + \frac{a^2}{c^2} k_2 V^{ic} + \left(\frac{c^3}{a^3} + k's\right) S^{ic} \\ y_5(c^-, \theta, \varphi) = \left(\frac{c^2}{a^2} + \frac{c}{a} k_s\right)(V^r + V) + \frac{ca}{b^2} k_1^s V^c + \left(1 + \frac{a}{c} k_2^s\right) V^{ic} + \left(1 + k_2^s\right) S^{ic} \\ \left[\frac{2}{15} \frac{b^2 \Omega^2}{GA^c q_0} \left[r^4 \left(y_6 - \frac{2}{r} y_5\right)\right]_c^b = h^c(V^r + V) + h_1^c V^c + h_2^c V^{ic} + h'_c S^{ic} \right. \end{cases} \quad (94)$$

G being the gravitational constant.

Perturbations on the products of inertia of different parts of the Earth may be computed in different ways. The inner core being assumed homogeneous, $\mathbf{C}^s$ is a function of the tesseral radial displacement of degree 2 at the ICB. For the stratified fluid core, [62] have shown that $\mathbf{C}^c$ are expressible in terms of y_6, y_5 at the CMB and the ICB. For the whole Earth, MacCullagh's theorem [51] implies that $\mathbf{C}$ is proportional to the tesseral surface potential of degree 2. Noting $y_i(a, \theta, \varphi) = y_i^c(a) Y_{21}^c + y_i^s(a) Y_{21}^s$ for $i = 1, \dots, 6$ and $y_6^{(c,s)}(r) = \frac{dy_6^{(c,s)}(r)}{dr} - 4\pi G \rho(r) y_1^{(c,s)}(r)$, we have

$$\begin{cases} C_{13} + iC_{23} = -\frac{a^3}{G} (y_5^c(a) + i y_5^s(a)) \\ C_{13}^c + iC_{23}^c = \frac{1}{5G} \left[r^4 (y_6^c(r) + i y_6^s(r)) - 2r^3 (y_5^c(r) + i y_5^s(r)) \right]_c^b \\ C_{13}^{ic} + iC_{23}^{ic} = -\frac{3}{2} \frac{A^{ic}}{c} (y_1^c(c) + i y_1^s(c)) \end{cases} \quad (95)$$

In complex form such as $W_2 = V_{21} + i\tilde{V}_{21}$ and $\theta^{ic} = \theta_1^{ic} + i\theta_2^{ic}$, and using the previous Love numbers,

$$\begin{cases} C_{13} + iC_{23} = \alpha A \left(\frac{k}{k_f} \frac{\omega}{\Omega} + \frac{k_1}{k_f} \frac{\omega^c}{\Omega} + \frac{k_2}{k_f} \frac{\omega^{ic}}{\Omega} - \frac{k}{k_f} \frac{3W_2}{\Omega^2 a^2} \right) \\ \quad - i\alpha^{ic} A^{ic} \Delta \rho \theta^{ic} \left(1 + \left(\frac{a}{c}\right)^3 k's \right) \\ C_{13}^c + iC_{23}^c = \frac{A^c q_0}{2} \left(h^c \frac{\omega}{\Omega} + h_1^c \frac{\omega^c}{\Omega} + h_2^c \frac{\omega^{ic}}{\Omega} - h^c \frac{3W_2}{\Omega^2 a^2} \right) \\ \quad + i\alpha^{ic} A^{ic} (1 - \Delta \rho) \theta^{ic} (1 + x_s h'^c) \\ C_{13}^{ic} + iC_{23}^{ic} = \frac{A^{ic} q_0}{2} \left(m \frac{\omega}{\Omega} + m_1 \frac{\omega^c}{\Omega} + m_2 \frac{\omega^{ic}}{\Omega} - m \frac{3W_2}{\Omega^2 a^2} \right) \\ \quad - i\alpha^{ic} A^{ic} \theta^{ic} (1 + x_s h'^s) \end{cases} \quad (96)$$

with $\Delta \rho = 1 - \rho^c(c)/\rho^{ic}$ and $x_s = \frac{4\pi G}{5} (\rho^{ic} - \rho^c(c))c/g_0$. $k_f = 3\alpha AG/\Omega^2 a^5$ is the secular Love number.

Gravitational and Pressure Torque

We come now to the gravitational and pressure torque on the inner core. In the initial state the core is in hydrostatic equilibrium: the equidensity, equipressure, equipotential surfaces are identical. Write $\alpha(r)$ for the hydrostatic

flattening within the Earth (computed from the Clairaut differential equation, using the PREM) and Ψ_{20} , Φ_{20} for the zonal centrifugal and gravitational potentials. We have

$$\Psi_{20}(r) + \Phi_{20}(r) = -\frac{2}{3}\alpha(r)g(r)r \quad (97)$$

where $g(r)$ is the radial gravity. The gravity on the surface will be denoted by $g_0 = g(a)$.

In a perturbed state, the mantle and the inner core are elastic. Let $\mathbf{n}^s$ be the outward normal on the inner core. The gravitational and pressure torque acting on the inner core (assumed homogeneous) is

$$\begin{aligned} \mathbf{\Gamma}^{ic} &= \int_{\text{inner core}} \mathbf{r} \wedge \rho^{ic} \nabla \left(\Phi^{ic}(r) + \frac{r^2}{a^2} V \right) dv - \int_{ICB} \mathbf{r} \wedge \mathbf{n}^s P^c ds \\ &= \int_{ICB} \mathbf{r} \wedge \mathbf{n}^s \left(\rho^{ic} \left(\Phi^{ic}(c) + \frac{c^2}{a^2} V \right) - P^c \right) ds \end{aligned} \quad (98)$$

where the gravitational potential Φ^{ic} at the ICB is

$$\Phi^{ic}(c) = y_5(c^-, \theta, \varphi) - V^{ic} - \frac{c^2}{a^2}(V^r + V) + \Phi_{20}(c)Y_{20} \quad (99)$$

The tangential gradient of pressure within the fluid core equals that of the product of density and the potential of the fluid; consequently at the ICB we have [46]

$$\begin{aligned} \int_{ICB} \mathbf{r} \wedge \mathbf{n}^s P^c ds &= \int_{ICB} \mathbf{r} \wedge \mathbf{n}^s \rho^c(c) \left[y_5(c^-, \theta, \varphi) - V^{ic} + \frac{c^2}{b^2} V^c \right. \\ &\quad \left. + (\Phi_{20}(c) + \Psi_{20}(c))Y_{20} \right] ds \\ &= \int_{ICB} \mathbf{r} \wedge \mathbf{n}^s \rho^c(c) \left(\Phi^{ic} + \frac{c^2}{a^2} V + \frac{c^2}{b^2} V^c + \frac{c^2}{a^2} V^r + \Psi_{20}(c)Y_{20} \right) ds \end{aligned} \quad (100)$$

Thus,

$$\begin{aligned} \mathbf{r} \wedge \mathbf{n}^s \left(\rho^{ic} \left(\Phi^{ic} + \frac{c^2}{a^2} V \right) - P^c \right) &= \mathbf{r} \wedge \mathbf{n}^s \rho^{ic} \left[\Delta \rho \left(\Phi^{ic} + \frac{c^2}{a^2} V + \Psi_{20}(c)Y_{20} \right) \right. \\ &\quad \left. - (1 - \Delta \rho) \left(\frac{c^2}{a^2} V^r + \frac{c^2}{b^2} V^c \right) - \Psi_{20}(c)Y_{20} \right] \end{aligned} \quad (101)$$

The two last terms correspond to the inertial torque.

Let

$$h(\theta, \varphi) = -\frac{2}{3}\alpha^{ic}cY_{20} + y_1(c, \theta, \varphi) \quad (102)$$

represent the ICB topography. The outward normal on the inner core is

$$\mathbf{n}^s = \begin{pmatrix} 1 \\ -\frac{\partial h}{\partial \theta} \\ 1 \\ -\frac{1}{\sin \theta} \frac{\partial h}{\partial \varphi} \end{pmatrix} \quad (103)$$

Introduce combinations of Love numbers:

$$\begin{cases} \xi^m = \frac{2\alpha^{ic}}{\Omega^2 c} (mg(c) - k_s g_0) \\ \xi^c = \frac{2\alpha^{ic}}{\Omega^2 c} (m_1 g(c) - k_1^s g_0) \\ \xi^{ic} = \frac{2\alpha^{ic}}{\Omega^2 c} (m_2 g(c) - k_2^s g_0) \end{cases} \quad ; \quad \begin{cases} \gamma^s = -2\alpha^{ic} \frac{g_0 x_s - g(c)}{\Omega^2 c} \\ \xi = x_s \frac{g_0 k_2'^s - g(c) h'^s}{g(c) - g_0 x_s} \end{cases} \quad (104)$$

$-c^2 \Omega^2 \gamma^s / 3$ is how much the mantle and the fluid core contribute to the zonal part of the potential within the inner core. For the PREM, $\gamma^s = 3.175851$.

Using all this notation, we have

$$\begin{aligned} I_1^{ic} + iI_2^{ic} = & -\alpha^{ic} A^{ic} \Omega^2 \gamma^s \Delta \rho \theta^{ic} (1 - \xi) - i \frac{g_0}{2} A^{ic} \Omega \Delta \rho (\xi^m \omega + \xi^c \omega^c + \xi^{ic} \omega^{ic}) \\ & + 3i A^{ic} \Delta \rho \frac{W_2}{a^2} \frac{g_0}{2} \xi^m - i(1 - \Delta \rho) A^{ic} \alpha^{ic} \Omega (\omega + \omega^c) \\ & + i \Omega^2 (C_{13}^{ic} + i C_{23}^{ic}) - 3i \alpha^{ic} A^{ic} \Delta \rho \frac{W_2}{a^2} \end{aligned}$$

Tidal Torque and Electromagnetic Coupling

The tesseral tidal potential of degree 2 acts on the equatorial bulge of the Earth and causes an equatorial torque

$$I_1^t + I_2^t = -\frac{3i\alpha A}{a^2} W_2 \quad (105)$$

Because of the conductivity of the lower mantle and of the inner core, a torque of magnetic friction appears at the CMB and the ICB. The equatorial magnetic torques T^c , T^{ic} acting on the fluid outer core and on the inner core have been computed from Buffett (1992):

$$\begin{aligned} T^c &= -A^c \Omega K_c (1 + i) \omega^c + A^{ic} \Omega K_{ic} (1 + i) \omega^{ic} \\ T^{ic} &= A^{ic} \Omega K_{ic} (1 + i) (\omega^c - \omega^{ic}) \end{aligned} \quad (106)$$

The frictional constants K_c and K_{ic} depend on the model of conductivity within the Earth and on the amplitude, as yet unknown, of the radial component of the geomagnetic field at the the CMB and the ICB.

Rotational Eigenmodes

We write the equatorial angular momentum equations (92) in the frequency (λ) domain:

$$(A) \begin{pmatrix} \omega \\ \omega^c \\ \omega^{ic} \end{pmatrix} = \begin{pmatrix} F \\ F^c \\ F^{ic} \end{pmatrix} \quad (107)$$

The matrix A has entries

$$\begin{aligned}
a_{11} &= i\lambda(1 + \alpha \frac{k}{k_f}) - i\alpha\Omega(1 - \frac{k}{k_f}) \\
a_{12} &= i(\lambda + \Omega)\left(\frac{A^c}{A} + \alpha \frac{k_1}{k_f}\right) \\
a_{13} &= i(\lambda + \Omega)\left(\frac{A^{ic}}{A} + \alpha \frac{k_2}{k_f} - \frac{\Omega}{\lambda} \alpha^{ic} \frac{A^{ic}}{A} \Delta\rho(1 + \frac{a^3}{c^3} k's)\right) \\
a_{21} &= i\lambda(1 + \frac{q_0}{2} h^c) \\
a_{22} &= i\lambda(1 + \frac{q_0}{2} h_1^c) + i\Omega(1 + \alpha^c + (1 - i)K_c) \\
a_{23} &= i\lambda \frac{q_0}{2} h_2^c + i\Omega \frac{A^{ic}}{A^c} (\alpha^{ic}(1 - \Delta\rho)(1 + x_s h'_c) - (1 - i)K_{ic}) \\
a_{31} &= i\lambda(1 + \frac{q_0}{2} m) - i\Omega \Delta\rho(\alpha^{ic} - \frac{q_0}{2} \xi^m) \\
a_{32} &= i\lambda \frac{q_0}{2} m_1 + i\Omega(\Delta\rho \frac{q_0}{2} \xi^c + \alpha^{ic}(1 - \Delta\rho) - (1 - i)K_{ic}) \\
a_{33} &= i\lambda(1 + \frac{q_0}{2} m_2) + i\Omega(1 + \frac{q_0}{2} \Delta\rho \xi^{ic} - \alpha^{ic}(1 + x_s h'^s) + (1 - i)K_{ic} \\
&\quad - \frac{\Omega}{\lambda} \alpha^{ic} \gamma^s \Delta\rho(1 - \xi))
\end{aligned} \tag{108}$$

and the excitation functions are

$$\begin{cases} F = 3i\alpha \frac{W_2}{a^2} \left(\frac{k}{k_f} \left(1 + \frac{\lambda}{\Omega}\right) - 1\right) \\ F^c = 3i \frac{q_0}{2} h^c \frac{W_2}{a^2} \frac{\lambda}{\Omega} \\ F^{ic} = 3i \frac{W_2}{a^2} \left(\Delta\rho \left(\frac{q_0}{2} \xi^m - \alpha^{ic}\right) + \frac{\lambda}{\Omega} \frac{q_0}{2} m\right) \end{cases} \tag{109}$$

The determinant of A vanishes for 4 eigenfrequencies: the Chandler wobble (motion of $\boldsymbol{\omega}$ with respect to the axis of the mantle) the inner core wobble (motion of $\boldsymbol{\omega} + \boldsymbol{\omega}^s$ with respect to the axis of the inner core), and two nearly diurnal modes, the Free Core Nutation (FCN) and the Free Inner Core Nutation (FICN). The analytical forms of the rotational eigenmodes are given in [24]. For the PREM, assuming no viscoelectromagnetic torque at the CMB and the ICB, we obtain the following elastic rotational eigenmodes:

$$\begin{aligned} \lambda_{FCN} &= -\Omega \left(1 + \frac{1}{458.6}\right), & \lambda_{FICN} &= -\Omega \left(1 - \frac{1}{484.9}\right), \\ \lambda_{CW} &= \frac{\Omega}{397.3}, & \lambda_{ICW} &= \frac{\Omega}{2319.3} \end{aligned} \tag{110}$$

The periods of the rotational eigenmodes depend strongly on flattenings. We presented here results for hydrostatic flattening, but heterogeneities in mantle density can create a non-negligible non-hydrostatic flattening for the ICB, the CMB, and Earth's surface.

How electromagnetic coupling on the core affects the FCN has been studied in detail. For different models of conductivity in the lower mantle (large conductivity of about $5 \times 10^5 \Omega\text{m}^{-1}$ within a 500 m-thick layer at the bottom of the lower mantle, or small conductivity of about $10 \Omega\text{m}^{-1}$ within a 2000 km-thick layer), it hardly perturbs the FCN period [5, 23] and it brings in a quality factor varying from 10^5 to 10^7 , i.e. large compared with the quality factor of the FCN observed using gravimeters and VLBI stacking which is about 31000 ± 11000 [15].

The influence on the FICN of electromagnetic torque acting on the elastic inner core and the associated perturbations in the observed nutations have

first been investigated by Buffett (1992), then by Buffett et al. (2002). From the analytical form of the frequency, we show that the FICN may be written, to first order in A^{ic}/A ,

$$\lambda_{FICN} = \lambda_{FICN}^{el} - \Omega K_{ic} + i\Omega K_{ic} \quad (111)$$

There damping with a quality factor $1/2K_{ic}$ and the FICN period is increased by the magnetic friction at the ICB. K_{ic} is proportional to the amplitude squared of B_r^{ICB} , radial component of the geomagnetic field at the ICB. This value is not precisely known. Assuming B_r^{ICB} is of the same order as B_r^{CMB} , the magnetic frictional constant K_{ic} is about 10^{-5} . With this kind of consideration, [47] estimated the strength of magnetic coupling at the CMB and the ICB from the amplitudes of the nutations obtained by analysis of VLBI data. For the real part of the magnetic torque they arrive at an estimate $K_{ic} = (100 \pm 8)10^{-5}$; it translates to a magnetic strength at the ICB of about 23 gauss, 5 times the strength at the CMB. The FICN period for such a frictional constant is about 920 ± 70 days.

The viscosity of the inner core can also perturb the period and the damping of the rotational eigenmodes of the Earth [24, 25].

Forced Solutions

We briefly discuss an approach to computing the rotational perturbation and deformation induced by the tidal gravitational potential. Write the tesseral tidal potential of degree 2 associated with a nearly diurnal tidal wave n_0 with frequency λ_x as [60]

$$V = \frac{3}{2}W_0(n_0) \sin 2\theta \sin(-\lambda_x t + \varphi) \quad (112)$$

$W_0(n_0)$ being the amplitude of the potential in m^2/s^2 given by Roosbeek (1995). Introducing

$$V = (V_{21} \cos \varphi + \tilde{V}_{21} \sin \varphi) 3 \cos \theta \sin \theta$$

and using the complex form, this nearly diurnal potential may be written

$$W_2(t) = V_{21} + i\tilde{V}_{21} = iW_0 e^{-i\Omega(1+x)t}$$

In the frequency domain $W_2(\lambda) = iW_0 \delta(\lambda + \Omega(1+x))$, where δ is the Dirac delta function. Next using the excitation functions F , F^c , F^{ic} defined in (109), we solve (107) for tidal waves in the frequency domain. Knowing ω , ω^c , ω^s , we compute the perturbation of gravity on the outer deformed surface and, after integration with respect to time, the nutations $\omega e^{i\Omega t}$. Details of this computation are found in [25].

Let us find ω for the precession: in the frequency domain we put $x = 0$ in the solution, then find

$$\omega = \frac{3\alpha W_0}{\Omega a^2} \quad (113)$$

The elasticity of the mantle and of the inner core as well as the existence of a fluid core do not affect the solution for the equatorial rotation perturbation of the Earth as a whole induced by the precession.

4.3 Discussion

Throughout this paper we have seen discrepancies between the rotation of a rigid planet and that of a more realistic planet which may be elastically deformed and made of a fluid core and a solid inner core. The elastic deformation is reflected in perturbations of the inertia tensor due to rotation itself and due to external sources: surface loading or tidal gravitational potential. We studied three angular velocities – of the mantle, as well as of the fluid core and of the solid inner core.

In a viscoelastic model of the Earth, it is further necessary to consider how viscosity perturbs the inertia tensor. We shall not treat this problem in any generality, but examine instead two simple instances that play an important role in the theory of rotation of a deformable body. The first is viscosity in the classical problem of secular deceleration under the action of the luni-solar frictional tidal torque; we will meet the concept of delayed rotational bulge. The second is large deviations of the axis of rotation over a geological timescale. The essential mechanisms are variations in the inertia tensor caused by an abrupt change in surface loading (deglaciation) or by convective dynamics within the mantle. In this problem, the relaxation of the rotational bulge will be crucial.

Secular Deceleration

It is known that 90% of the deceleration of the Earth comes from dissipation by oceans (e.g. [39]). So our discussion here will be more theoretical and realistic. For a viscoelastic Earth subject to a luni-solar tidal potential, the angular momentum theorem says

$$\frac{d\mathbf{H}}{dt} + \boldsymbol{\omega} \wedge \mathbf{H} = \mathbf{\Gamma}^t \quad (114)$$

where $\mathbf{\Gamma}^t$ is the tidal torque. In the reference frame of the mantle, let $\mathbf{u}$ be the direction of the body responsible for the tides, M , r its mass and distance. The torques are then

$$\begin{cases} \Gamma_1^t = \frac{3GM}{r^3} (-u_1 u_3 C_{12} + u_1 u_2 C_{13} + u_2 u_3 (C_{33} - C_{22}) - C_{23} (u_3^2 - u_2^2)) \\ \Gamma_2^t = \frac{3GM}{r^3} (-u_2 u_1 C_{23} + u_2 u_3 C_{21} + u_3 u_1 (C_{11} - C_{33}) - C_{31} (u_1^2 - u_3^2)) \\ \Gamma_3^t = \frac{3GM}{r^3} (-u_3 u_2 C_{31} + u_3 u_1 C_{32} + u_1 u_2 (C_{22} - C_{11}) - C_{12} (u_2^2 - u_1^2)) \end{cases} \quad (115)$$

In these formulae it is usual to look at only the third terms, in differences of principal moments of inertia, in I_1^t and I_2^t when studying nutation, and in I_3^t when studying deceleration. We shall, however, look at all the terms so as to keep the symmetry. Perturbations on the inertia tensor in these formulae (115) result from viscous deformations of the planet. They are computed from MacCullagh's theorem. Take a rheology where the Love numbers are frequency-dependent. We have in the temporal domain

$$C_{13}(t) = -\frac{a^5}{G}k_2(t) * \left(\frac{GM}{r^3}u_1(t)u_3(t) \right) \quad (116)$$

Note that the distance r too is time-dependent. In this temporal domain the remaining components are

$$\begin{cases} C_{23}(t) = -\frac{a^5}{G}k_2(t) * \left(\frac{GM}{r^3}u_2(t)u_3(t) \right) \\ C_{12}(t) = -\frac{a^5}{G}k_2(t) * \left(\frac{GM}{r^3}u_1(t)u_2(t) \right) \\ C_{22}(t) - C_{11}(t) = -\frac{a^5}{G}k_2(t) * \left(\frac{GM}{r^3}(u_1^2(t) - u_2^2(t)) \right) \\ C_{33}(t) - C_{22}(t) = -\frac{a^5}{G}k_2(t) * \left(\frac{GM}{r^3}(u_2^2(t) - u_3^2(t)) \right) \\ C_{11}(t) - C_{33}(t) = -\frac{a^5}{G}k_2(t) * \left(\frac{GM}{r^3}(u_3^2(t) - u_1^2(t)) \right) \end{cases} \quad (117)$$

Substitute these expressions for inertia perturbations into the system (115). The resulting formulae are symmetric. Note that these formulae depend only on the tidal Love number $k_2(t)$, whatever the rheology. The phase delay is consequently given by Q_{k_2} as in (73). The energetic factor Q_E and the rheological factor Q_μ are irrelevant.

The simplicity of this problem in terms of the tidal Love number k_2 is as striking as its complexity involved in the choice of rheology.

Large Deviations of Axis of Rotation

Here is a final problem that arises in the rotation of a deformable planet, whereas it does not arise at all for a rigid planet: large deviations of the axis of rotation, also known as polar wander. As seen in a previous section, the rotation of the Earth together with its internal fluid and solid cores is governed by the angular momentum theorem written in a frame co-rotating with the mantle. Approximations in these equations are made to solve them, following the time-scale of each excitation mechanism and following the rheology of the mantle:

- Short time-scale: the Earth has an elastic behavior and ω remains close to the initial angular velocity Ω (sidereal rotation); the linearized Liouville equations (with the four classical rotational eigenmodes) model changes in the rotation due to the core/mantle or atmosphere/mantle coupling.

- Time-scale of a thousand years: the Earth has a viscoelastic behavior and the linearized Euler equations model changes in the rotation due to the last deglaciation.
- Time-scale of a million years: the Earth has a viscous behavior and the nonlinear equations may be used to estimate, for example, the effects of the mantle convection on the polar wander.

The current observation of the secular polar wander points to a superposition of the effects of the post-glacial rebound plus a long-term trend resulting from the convective mantle mass anomalies.

4.4 Conclusion

What have we learnt in this paper?

The simplest model of the Earth is that of a rigid body, whose dynamics of rotation is perfectly understood. We try to have a more realistic model by including deformability, and deformability produces new phenomena in the rotation of the Earth; conversely, rotation causes variations in the inertia tensor of the Earth. These deformations may be elastic or viscous, depending on the time-scale of the sources of excitation. As deformations are small and the Earth is in quasi-hydrostatic equilibrium, our analysis proceeds via perturbation theory.

Two classical problems come from elasticity: the problem of transition from the Eulerian to Chandlerian motion as well as driving of this motion by surface loading, and the problem of variations in the axial rotation induced by zonal surface loading and luni-solar tides. Two less classical problems come from viscosity: the problem of secular deceleration by the frictional tidal torque, and the problem of large deviations of the axis of rotation resulting from the last deglaciation or from the mantle convection. The fact that the Earth has internal fluid parts creates additional eigenmodes of rotation, and the fluid core modifies the Chandler period and perturbs the amplitude and phase of some nutations.

Where do we go from here?

Current researches on the rotation of a deformable Earth focus on problems of dissipation by magnetic coupling across the core/mantle boundary, and more generally on problems of coupling between the rotation and the electromagnetic torsional oscillations. But the current theory is linear. The next challenge is to develop a nonlinear theory that encompasses both the rotation in the presence of deformability and various dissipative processes.

References

1. D'Alembert, J., 1749. Recherche sur la précession des equinoxes et sur la nutation de l'axe de la Terre, dans le système newtonien. Ed. David l'Ainé, Paris.

2. Alterman, Z., Jarosch, H., and Pekeris, C.H., 1959. Oscillation of the Earth. *Proc. R. Soc. Lond. A* **252**, 80–95.
3. Anderson, D.L., 1982. The Earth's core and the phase diagram of iron. *Phil. Trans. R. Soc. Lond. A* **306**, 21–35.
4. Anderson, D.L., Isaak, D., and Oda, H., 1992. High-temperature elastic constant data on minerals relevant to geophysics. *Rev. Geophys.* **30**, 1, 57–90.
5. Buffett, B., 1992. Constraints on magnetic energy and mantle conductivity from the forced nutations of the Earth. *J. Geophys. Res.* **97**, B13, 19581–19597.
6. Buffett, B., Mathews, P.M., and Herring, T.A., 2002. Modeling of nutation and precession: effects of electromagnetic coupling. *J. Geophys. Res.* **107**, 4, 15 pp.
7. Bullen, K.E., 1942. The density variation of the Earth's central core. *Bull. Seismol. Soc. Amer.* **32**, 19–29.
8. Bullen, K.E., 1975. The Earth's density. Chapman and Hall, London.
9. Cameron A.G.W., 1988. Origin of the solar system. *Ann. Rev. Astron. Astrophys.* **26**, 441–472.
10. Chandler, S.C., 1891. On the variation of the latitude. *Astron. J.* no. 248 and 249.
11. Coirier, J., 2001. Mécanique des milieux continus. Dunod, Paris.
12. Dahlen, F.A. and Tromp, J., 1998. Theoretical Global Seismology. Princeton University Press.
13. Darwin, G.H., 1883. Attempted evaluation of the rigidity of the Earth from the tides of long period. *Scientific Papers I* **9**, 340–346.
14. Darwin, G.H., 1878. On the influence of geological changes on the Earth's axis of rotation. *Phil. Trans. R. Soc. Lond.* **167**, part I, 271–313; also 1879. **170**, 1–35; 1880. **170**, 497–530.
15. Defraigne, P., Dehant, V., and Hinderer, J., 1995. Correction to 'Stacking gravity tide measurements and nutation observations in order to determine the complex eigenfrequency of the Nearly Diurnal Free Wobble', *J. Geophys. Res.* **100**, 2041–2042.
16. Deparis, V. and Legros, H., 2000. Voyage à l'intérieur de la Terre. CNRS éd.
17. Dziewonski, A.M. and Anderson, D.L., 1981. Preliminary Reference Earth Model PREM, *Phys. Earth Planet. Int.* **25**, 297–356.
18. Euler, L., 1749. Recherche sur la précession des équinoxes et sur la nutation de l'axe de la Terre. *Mém. Acad. R. Sci. Berlin*; reprinted in 'Commentationes astronomicae', Societatis Scientiarum Naturalium Helveticae, Leo Courvoisier, 1941.
19. Euler, L., 1758. Du mouvement de rotation des corps solides autour d'un axe variable. *Mém. Acad. R. Sci. Berlin*.
20. Farrell, W.E., 1972. Deformation of the Earth by surface loads. *Rev. Geophys. and Space Phys.* **10**, 3, 761–797.
21. Gegout, P., 1995. De la variabilité de la rotation de la Terre et du champ de gravité, conséquence aux dynamiques de l'atmosphère et des océans. Thèse, Strasbourg.
22. Gold, T., 1955. Instability of the Earth's axis of rotation. *Nature* **175**, 526–529.
23. Greff-Lefftz, M. and Legros, H., 1999. Magnetic field and rotational eigenfrequencies. *Phys. Earth Planet. Int.* **112**, 21–41.
24. Greff-Lefftz, M., Legros, H., and Dehant, V., 2000. Influence of inner core viscosity on the rotational eigenmodes of the Earth. *Phys. Earth Planet. Int.* **122**, 187–203.

25. Greff-Lefftz, M., Dehant, V., and Legros, H., 2002. Effects of inner core viscosity on gravity changes and spatial nutations induced by luni-solar tides. *Phys. Earth Planet. Int.* **129**, 31–41.
26. Gutenberg, B., 1914. Über Erdbebenwellen VII. *Nach. Ges. Wiss. Göttingen Math. Phys. Klasse A*, 166–218.
27. Haskell, N.A., 1935. The motion of a viscous fluid under a surface load, *Physics* **6**, 265 pp.
28. Heiskanen, W. and Moritz, H., 1967. Physical Geodesy. W.H. Freeman, San Francisco-London.
29. Hess, H.H., 1962. History of ocean basin. In A.E.J. Engel Petrologic studies, Geological Society of America, Boulder.
30. Hopkins, W., 1839. Researches in physical geology. On the phenomena of precession and nutation, assuming the fluidity of the interior of the Earth. *Phil. Trans. R. Soc. Lond.* **129**, 381–423; also **130**, 193–208.
31. Hopkins, W., 1847. Geological theories of elevation of earthquakes. Report of the British Association for the Advancement of Science for 1847.
32. Hough, S.S., 1895. The oscillations of a rotating ellipsoidal shell containing fluid. *Phil. Trans. R. Soc. Lond.* **186**, 469–505.
33. Inglis, D.R., 1957. Shifting of the Earth's axis of rotation. *Rev. Modern Phys.* **29**, 9–19.
34. Jameson, T.F., 1865. On the history of the last geological changes in Scotland. *Quart. J. Geol. Soc. Lond.* **21**, 178 pp.
35. Jeffreys, H., 1926. The viscosity of the Earth. *Month. Not. R. Astr. Soc.* **1**, 84–112.
36. Jeffreys, H., 1930. The resonance theory of the origin of the Moon. *Month. Not. R. Astr. Soc.* **91**, 169–173.
37. Jeffreys, H., 1951. The Earth (3rd ed.). Cambridge University Press.
38. Jeffreys, H. and Vicente, R.O., 1957. The theory of nutation and the variation of latitude. *Month. Not. R. Astr. Soc.* **117**, 142–173.
39. Lambeck, K., 1980. The Earth's variable rotation. Cambridge University Press.
40. Lehmann, I., 1936. Publication du Bureau Central Seismologique International; Travaux scientifiques, 14, 87–115.
41. Le Pichon, X., 1968. Sea-floor spreading and continental drift. *J. Geophys. Res.* **73**, 3661–3697.
42. Liu, L.G., 1976. The high-pressure phase of $MgSiO_3$. *Earth Planet. Sci. Lett.* **31**, 200–208.
43. Love, A.E.H., 1909. The yielding of the Earth to disturbing forces. *Proc. R. Soc. Lond. A* **82**, 73–88.
44. Love, A.E.H., 1911. Some Problems of Geodynamics. Cambridge University Press.
45. McKenzie, D.P. and Parker, R.L., 1967. The North Pacific: an example of tectonics on a sphere. *Nature* **216**, 1276–1279.
46. Mathews, P.M., Buffett, B.A., Herring, T.A. & Shapiro, I.I., 1991. Forced nutations of the Earth: influence of inner core dynamics: I. Theory. *J. Geophys. Res.* **96** B5, 8219–8242.
47. Mathews, P.M., Herring, T.A., and Buffett, B.A., 2002. Modeling of nutation-precession: new nutation series for nonrigid Earth and insights into the Earth's interior. *J. Geophys. Res.* **107**, 4, 30 pp.
48. Milankovitch, M., 1934. Der Mechanismus des Polverlagerungen und die daraus sich ergebenden Polbahnkurven. *Gerlands Beitr. Geophys.* **42**, 70–97.

49. Molodensky, M.S., 1961. The theory of nutation and diurnal Earth tides. *Commun. Observ. Roy. Belg.* **188**, 25–56.
50. Moritz, H., 1990. The Figure of the Earth. Wichmann.
51. Munk, W.H. and MacDonald, G.J.F., 1960. The Rotation of the Earth, Cambridge University Press.
52. Newcomb, S., 1892. On the dynamics of the Earth's rotation with respect to periodic variations of latitude. *Month. Not. R. Astr. Soc.* **52**, 336–341.
53. Newton, I., 1687. Principes mathématiques de la philosophie naturelle. Traduit en français par La Marquise du Chastelet, 1756; reprinted A. Blanchard, 1966, Paris.
54. O'Connell, R.J. and Budiansky, B., 1978. Measures of dissipation in visco-elastic media. *Geophys. Res. Lett.* **5**, 5–8.
55. Patterson, Cl., 1956. Age of meteorites and the Earth. *Geochimica et Cosmochimica Acta* **10**, 230–237.
56. Peltier, W.R., 1989. Mantle Convection, Plate Tectonics and Global Dynamics. Gordon and Breach.
57. Peltier, W.R., 1974. Impulse response of a Maxwell Earth. *Rev. Geophys. Space Phys.* **12**, 649–669.
58. Poirier, J.P., 1991. Introduction to the Physics of the Earth Interior. Cambridge University Press.
59. Poincaré, H., 1910. Sur la précession des corps déformables. *Bull. Astr.* **27**, 321–356.
60. Roosbeek, F., 1995. RATGP95: a harmonic development of the tide-generating potential using an analytical method. *Geophys. J. Int. Astr. Soc.* **126**, 197–204.
61. Safronov, V.S., 1969. Evolution of the protoplanetary cloud and formation of the Earth and planets. Translated by the Israel Program for Scientific Translation, 1972.
62. Sasao, T., Okubo, S., and Saito, M., 1980. A simple theory on the dynamical effects of a stratified fluid core upon the nutational motion of the Earth, *Proc. IAU Symp. 78, 'Nutation and the Earth's rotation'*, 165–183, eds Fedorov, E.P., Smith, M.L., and Bender, P.L., Reidel, Dordrecht.
63. Schiaparelli, I.V., 1889. De la rotation de la Terre sous l'influence des actions géologiques. St Petersburg, Imprimerie de l'Académie Impériale des Sciences. Vass. Ostr., 9e ligne, no. 12.
64. Schubert, G., Turcotte, D.L. and Olson P., 2001. Mantle Convection in the Earth and Planets. Cambridge University Press.
65. Smith, M., 1974. The scalar equations of infinitesimal elastic-gravitational motion for a rotating, slightly elliptical Earth, *Geophys. J. Int. Astr. Soc.* **37**, 491–526.
66. Stacey, F.D., 1992. Physics of the Earth (3rd ed.). Brookfield Press.
67. Suess, E., La face de la Terre. Traduit en français par E. de la Margerie. Armand Colin, 1897–1918.
68. Takeuchi, H., 1950. On the tides of the compressible Earth of variable density and elasticity. *Trans. Amer. Geophys. Union* **31**, no. 5, 651–689.
69. Thomson, Sir W., 1862. Dynamical problems regarding elastic spheroidal shells and spheroids of incompressible liquid. *Phil. Trans. R. Soc. Lond.* **153**, 583–608.
70. Thomson, Sir W. and Tait, P.G., 1879. Treatise on Natural Philosophy. Cambridge University Press.
71. Wahr, J.M., 1981. Body tides on the elliptical, rotating, elastic and oceanless Earth. *Geophys. J. R. Astr. Soc.* **64**, 677–703.

- 72. Wetherill G.W., 1980. Formation of the terrestrial planets. *Ann. Rev. Astron. Astrophys.* **18**, 77–113.
- 73. Williamson, E.D. and Adams, L.H., 1923. Density distribution in the Earth. *J. Wash. Acad. Sci.* **13**, 413–428.
- 74. Yuen, D.A. and Peltier, W.R., 1983. Normal modes of the viscoelastic Earth. *Geophys. J. R. Astr. Soc.* **69**, 495–526.
- 75. Zschau, J., 1978. Tidal friction in the solid Earth. In ‘Tidal Friction and the Earth’s rotation’, eds. Brosche, P. and Sunderman, J., Springer-Verlag, Berlin.

Modelling and Characterizing the Earth's Gravity Field: From Basic Principles to Current Purposes

Florent Deleflie and Pierre Exertier

OCA/GEMINI Av. N. Copernic, F-06130 Grasse

Florent.Deleflie@obs-azur.fr

Pierre.Exertier@obs-azur.fr

Abstract. The Earth is a specific body in the Universe. Nowadays, thanks to the high accuracy of the measurements obtained by different techniques of geodesy, its shape, as well as its gravity field, can be described and characterized in a very precise way, with thousands of parameters: the Earth is a spherical body only in a very first approximation!

In this paper, we start from Newton's universal law of gravitation; we remind the reader of some basic principles of the field theory, in order to get the development of a gravity field in spherical harmonics. Then we show the usual method of orbital dynamics used to get the coefficients of such a development.

In practice, the development in spherical harmonics can also be used to describe the gravity field of each celestial body which does not differ too much from a sphere. But, it is all the more well adapted to the Earth as the temporal variations of many coefficients, which can be very high, can be linked to global geophysical and geodynamical phenomena, enlightening the activity of an "alive planet". These phenomena, linked more or less to mass transfers between the different components of the Earth – solid earth, oceans, atmosphere in particular –, are briefly discussed.

We conclude with the current purposes about the modelling of the Earth gravity field, and with the required links which have to be kept and developed between geophysics, space geodesy, and celestial mechanics.

1 Introduction

In a dictionary, a "field" is defined as a limited area in space which has some characteristic properties. That means that a field is rather described by the way it interacts with its neighbourhood, than by its source itself, which is the case for an electric field, a magnetic field, or a gravity field. For a given field, the motion of a test particle of coordinates denoted (x, y, z) is affected, whatever the interaction law between 2 particles, by the charged particles of coordinates denoted (x_P, y_P, z_P) constituting the field. For a gravity field, the interaction law is of course Newton's universal law of gravitation.

In this paper, we remind the reader of the representation of the gravity field of a non spherical body usually used in space geodesy, as well as the way to compute, from measurements, the different parameters involved in that representation. We mention the level of accuracy reached for these parameters thanks to new measurements dedicated to the study of the gravity field of the Earth, and we comment the gravity field models recently computed.

Many books and articles detail these topics (see [16] for basic principles, and [2] for the most recent developments). Here, we aim at connecting the representation of a gravity field, the way to characterize it from the equations of dynamics of an artificial satellite, and the relationships with the shape of the Earth. As a consequence, we aim at presenting the way astronomers and geodesists can help, from space techniques, geophysicists to force some geophysical models in terms, for example, of global mass transfers.

2 Basic Principles

In this section, we give the general results which can be deduced from the universal law of gravitation. We start with the definition of the gravity potential generated by a punctual body of mass M_E . But, since the Earth is not an homogeneous sphere, this definition has to be generalized to a gravity field generated by elementary particles of mass dM . Let us note that the structure of the Earth is affected by many phenomena which modify the distribution of the mass (because of tides, for example); as a consequence, the gravity potential is time dependent.

2.1 Mass and Gravitation

The notion of gravity field, containing at the surface of a body both gravity and centrifugal forces, is closely linked with the notion of “shape” of that body. The surface gravity acceleration can directly be measured using a gravimeter. But, the observations of local gravity at any geometrical point at the surface of a planet are delicate to carry out: it is complicated to interpret given environmental and topographic settings. Artificial satellites which have been launched since half a century provide a new tool for improved activities in geodesy, through the dynamical description of their environment they give: it is true for the Earth, obviously, but also for Mars, Venus, and the Moon, in particular.

According to Newton’s universal law of gravitation, the force $\vec{F}$ of attraction between 2 particles denotes C and S of masses M_E and m_s respectively, at a distance r from each other, is:

$$\|\vec{F}\| = \frac{GM_E m_s}{r^2} \quad (1)$$

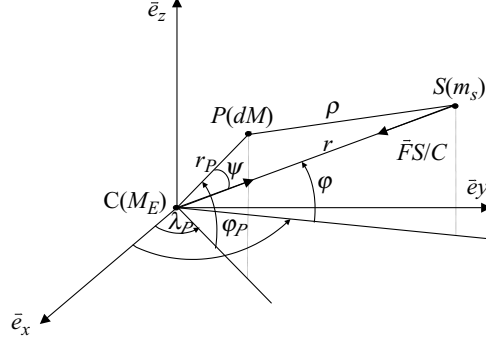


Fig. 1. The Newton law. When the whole mass of the Earth M_E is concentrated in C , the gravitational force acting on S is $\vec{F} = \vec{F}_{C/S}$. In the general case, the elementary masses dM of all the points P constituting the body have to be taken into account. S and P points are located thanks to their spherical coordinates (r, λ, φ) and $(r_P, \lambda_P, \varphi_P)$, respectively in an arbitrary frame $(C, \vec{e}_x, \vec{e}_y, \vec{e}_z)$ centred in C . The quantities ρ and ψ are also defined by: $\rho = ||\vec{PS}||$, $\psi = (\vec{CS}, \vec{CP})$

where G is the gravitational constant. Hereafter, $\vec{F}$ denotes the force created by C on S : $\vec{F} = \vec{F}_{C/S}$ (Fig. 1).

The coordinates of the point S are denoted (x, y, z) in an arbitrary frame $(C, \vec{e}_x, \vec{e}_y, \vec{e}_z)$ centred in C . The components of $\vec{F}$ are therefore:

$$\vec{F} = \begin{pmatrix} -\frac{\mu m_s x}{r^3} \\ -\frac{\mu m_s y}{r^3} \\ -\frac{\mu m_s z}{r^3} \end{pmatrix} \quad (2)$$

where $\mu = GM_E$. The corresponding force function U , or *potential*, is defined by $\vec{F} = \vec{\text{grad}} U$. It is defined as a positive quantity, which is consistent with the sign convention of geodesy¹. We get the newtonian potential

$$U = \frac{\mu m_s}{r} \quad (3)$$

The main advantage of using such a potential is that the whole reasoning can be done with one single scalar quantity. For example, adding together the second derivatives for the three coordinates, we get Laplace's equation (outside the Earth) [26]:

$$\frac{\partial^2 U}{\partial x^2} + \frac{\partial^2 U}{\partial y^2} + \frac{\partial^2 U}{\partial z^2} = \Delta U = 0 \quad (4)$$

¹in most of other fields in physics, a potential is conventionally taken to be negative.

If $\vec{a}$ denotes the acceleration of the point S , the integration of the equations of dynamics $\vec{F} = m_s \vec{a}$ leads to the definition of the well-known keplerian elements. It is important to note that the mass m_s can be simplified in both sides of these equations. As a consequence, all the following potentials are expressed per unit of mass. Moreover, let us note that the equipotential surfaces $U = U_0$ corresponding to (3) can easily be linked to the shape of the body producing such a potential, namely a homogeneous sphere of constant density. Even in that example, the knowledge of the potential does not determine the radius of that sphere in an unique way: another criterion, detailed hereafter, can be chosen to define the “mean” radius of the Earth.

2.2 Potential Generated by a Continuous Body

The potential corresponding to the real spatial distribution of mass at the surface of or inside the Earth is based on the previous result: the potential generated by several particles is the sum of the potentials they generate individually. If these particles are conglomerated to form a continuous body of variable density σ , this summation can be replaced by an integration over the volume of the body (each elementary volume $dx_P dy_P dz_P$ has an elementary mass $dM = \sigma(x_P, y_P, z_P) dx_P dy_P dz_P$):

$$U = \iiint_V \frac{G\sigma(x_P, y_P, z_P)}{\rho} dx_P dy_P dz_P \quad (5)$$

Let us note that the potential U is a function of the coordinates of the point S , and that in particular ρ can be expressed thanks to the rectangular (x, y, z) or spherical (r, λ, φ) coordinates of S (Fig. 1).

The distribution can also be located on a surface Σ , with a density denoted δ . In that case, the corresponding potential is expressed with a double integral:

$$U = \iint_{\Sigma} \frac{G\delta(u, v)}{\rho} du dv \quad (6)$$

As a consequence of the mass distribution into and onto the Earth (solid earth, ocean, atmosphere), which include their time dependant variations, the goal of geodesy and space geodesy consists in characterizing and studying the local and global aspects, respectively. The study of the motion of artificial satellites reflects the global deformations of the central body. Seen as sensors of forces, they can enlighten from space the global deformations of the central body (see hereafter).

2.3 Potential Generated by a Continuous Body in Rotation

It is in fact links between the gravity field and the shape of a planet which are at stake, and in particular the shape of the Earth.

As equipotentials $U = U_0$ of (3) characterize the shape of the central spherical homogeneous body creating U , the equipotentials of the complete gravity field characterize the shape of the Earth. In practice, the potential of rotation has also to be taken into account: these two potentials combine to make what is called the potential of gravity at the surface of the Earth

$$W = U + \frac{1}{2}\omega^2 u^2 \cos^2 \varphi$$

where ω is the rate of rotation, and u the distance of a point at the surface of the Earth from the rotation axis. As a consequence, equipotentials $W = W_0$ characterize the shape of the Earth. The value of W_0 is chosen such as representing, in the best way, the mean sea level, called the *geoid* (Fig. 2): on an equipotential, the gravity is normal with respect to the geoid, and water does not flow. The shape of the Earth is approximated within about 10^{-5} of the radius vector by an ellipsoid of revolution. The shape of this ellipsoid is conventionally defined by the flattening $f = \frac{R_0 - R_p}{R_0}$, where R_0 is the equatorial radius, and R_p the polar radius. The flattening can of course be expressed with respect to ω .

The reference of the geoid from an ellipsoid is historical, but is used to see the gravity anomalies (due to mountains for example) as small quantities.

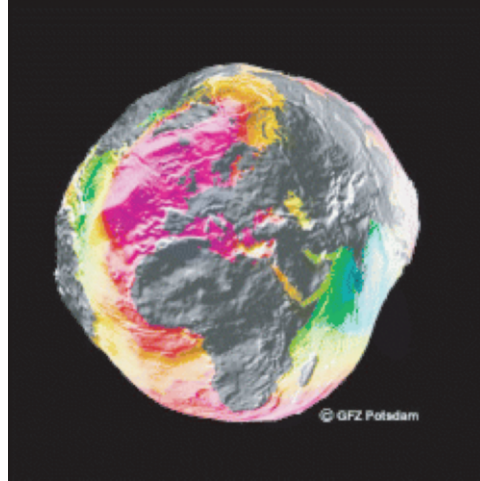


Fig. 2. The geoid. At a given time, the geoid is the most graphical representation of the Earth gravity field. The visual deformations, which are exaggerated in this figure, have to be referred to a reference ellipsoid (with an equatorial radius R_0 and with a flattening f)

The ellipsoid has the same centre of mass, the same rate of rotation, and the same mass as the Earth.

3 Coefficients Characterizing the Gravity Field

The first step to characterize the gravity field of a continuous body, through the corresponding potential, consists in describing it through several coefficients which depend only on the distribution of mass. Formulae (5) and (6) enclose also the coordinates of the sensor of forces, and have therefore to be modified.

It is easy to check that (5) and (6) verify the Laplace (4), since the point (x, y, z) is assumed to be located outside the body (on the contrary, inside, these potentials verify the Poisson equation [26]). The Laplace equation can be seen as a partial derivatives equation. Solving it permits to get an explicit formulation of U with respect to the coordinates of the point S . The other coefficients involved in that formulation are precisely those characterizing the mass distribution of the central body.

3.1 Legendre Polynomials

Applying the Al Kashi relation in the triangle CPS (Fig. 1) gives:

$$\rho^2 = r_P^2 + r^2 - 2rr_P \cos \psi$$

$\frac{r_P}{r}$ being small compared to 1, $\frac{1}{\rho}$ can be developed in a power series of $\frac{r_P}{r}$:

$$\frac{1}{\rho} = \frac{1}{r} \sum_{n \geq 0} \left(\frac{r_P}{r} \right)^n P_n(\cos \psi) \quad (7)$$

P_n are called Legendre polynomials. An explicit formulation is given in [26]. Defining

$$U_n = \frac{G}{r} \iiint_{\mathcal{V}} \left(\frac{r_P}{r} \right)^n P_n(\cos \psi) dM \quad (8)$$

we have $U = \sum_{n \geq 0} U_n$. The partial derivatives equation verified by U_n , expressed in spherical coordinates, reads:

$$\frac{\partial}{\partial r} \left(r^2 \frac{\partial U_n}{\partial r} \right) + \frac{1}{\cos \varphi} \frac{\partial}{\partial \varphi} \left(\cos \varphi \frac{\partial U_n}{\partial \varphi} \right) + \frac{1}{\cos^2 \varphi} \frac{\partial^2 U_n}{\partial \lambda^2} = 0 \quad (9)$$

3.2 Spherical Harmonics

Defining now $S_n(\lambda, \varphi) = r^{n+1}U_n$, we finally get the partial derivative equation to be solved:

$$\frac{1}{\cos^2 \varphi} \frac{\partial^2 S_n}{\partial \lambda^2} + \frac{1}{\cos \varphi} \frac{\partial}{\partial \varphi} \left(\cos \varphi \frac{\partial S_n}{\partial \varphi} \right) + n(n+1)S_n = 0 \quad (10)$$

The resolution of (10) leads to the definition of the so-called Legendre associated functions, which depend on a pair of subscripts n, m , and which permit, moreover, to formulate the dependency of $P_n(\cos \psi)$ with respect to λ and φ [26]:

$$P_n(\cos \psi) = \sum_{m=0}^n \delta_m \frac{(n-m)!}{(n+m)!} P_{n,m}(\sin \varphi) P_{n,m}(\sin \varphi_P) \cos m(\lambda - \lambda_P) \quad (11)$$

with:

$$\begin{cases} \delta_m = 1 & \text{if } m = 0 \\ \delta_m = 2 & \text{if } m \geq 1 \end{cases}$$

Equation (8) can, hence, be expanded:

$$\begin{aligned} S_n(\lambda, \varphi) &= GM_E P_n(\sin \varphi) \iiint_V r_P^n P_n(\sin \varphi_P) \frac{dM}{M_E} \\ &+ 2GM_E \sum_{m=1}^n \frac{(n-m)!}{(n+m)!} P_{n,m}(\sin \varphi) \cos(m\lambda) \iiint_V r_P^n P_{n,m}(\sin \varphi_P) \cos(m\lambda_P) \frac{dM}{M_E} \\ &+ 2GM_E \sum_{m=1}^n \frac{(n-m)!}{(n+m)!} P_{n,m}(\sin \varphi) \sin(m\lambda) \iiint_V r_P^n P_{n,m}(\sin \varphi_P) \sin(m\lambda_P) \frac{dM}{M_E} \end{aligned} \quad (12)$$

It is therefore possible to define $Y_{n,m}(\lambda, \varphi)$, called spherical harmonics of degree n and order m :

$$S_n(\lambda, \varphi) = \sum_{m=0}^n Y_{n,m}(\lambda, \varphi) \quad (13)$$

Referring to Eq. (12), each spherical harmonics contains a cosinus and a sinus term. Figure 3 shows the shape of the first spherical harmonics.

3.3 Development of the Gravity Field in Spherical Harmonics

The quantity

$$\iiint_V r_P^n P_n(\sin \varphi_P) \frac{dM}{M_E}$$

being a characteristic of the Earth, it is possible to define:

$$I_n = \iiint_V r_P^n P_n(\sin \varphi_P) \frac{dM}{M_E} \quad (14)$$

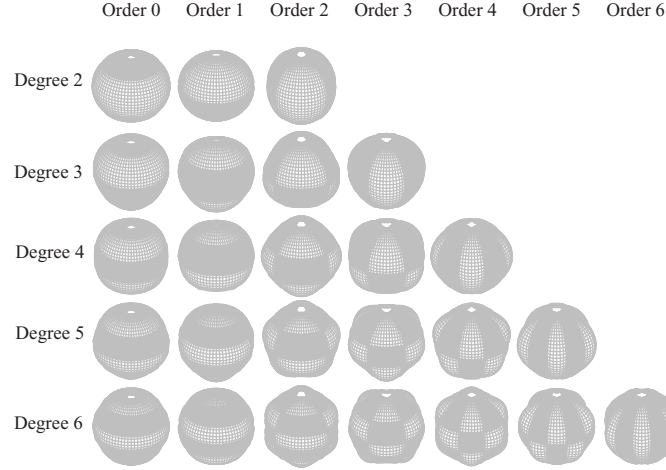


Fig. 3. Spherical harmonics (cosinus terms). Basic spherical functions developed up to degree and order 6. The shape of the Earth is modelled by a weighted sum of these basic functions

Moreover, this integral having the dimension of a length to the power n , coefficients J_n , without any dimension, are preferred to the coefficients I_n :

$$I_n = -J_n(R_0)^n$$

where R_0 is a conventional mean equatorial radius of the Earth², and where the sign '-' makes the coefficient J_2 be positive. These coefficients do not depend on the longitude ($m = 0$). Those depending on the longitude, denoted $C_{n,m}$ and $S_{n,n}$, are defined for $m \geq 1$ in the same way:

$$\begin{aligned} (R_0)^n C_{n,m} &= 2 \frac{(n-m)!}{(n+m)!} \iiint_{\mathcal{V}} r_P^n P_{n,m}(\sin \varphi_P) \cos(m\lambda_P) \frac{dM}{M_E} \\ (R_0)^n S_{n,m} &= 2 \frac{(n-m)!}{(n+m)!} \iiint_{\mathcal{V}} r_P^n P_{n,m}(\sin \varphi_P) \sin(m\lambda_P) \frac{dM}{M_E} \end{aligned} \quad (15)$$

As a result, the potential existing in (r, λ, φ) , or in $(x = r \cos \varphi \cos \lambda, y = r \cos \varphi \sin \lambda, z = r \sin \varphi)$, reads:

$$U(r, \lambda, \varphi) = \frac{\mu}{r} + \frac{\mu}{r} \sum_{n \geq 1} \sum_{m=0}^n \left(\frac{R_0}{r} \right)^n \left(C_{n,m} \cos(m\lambda) + S_{n,m} \sin(m\lambda) \right) P_{n,m}(\sin \varphi) \quad (16)$$

²any other quantity with the dimension of a length could also have been chosen.

where the definitions of $C_{n,m}$ and $S_{n,m}$ have been extended to $m = 0$, knowing that $P_{n,0} = P_n$: $C_{n,0} = -J_n$, and $S_{n,0} = 0$. The dependency of (16) can be transformed to other sets of variables, if necessary (in keplerian orbital elements for example [20]).

An important property of the surface spherical harmonics $S_{n,m}(\lambda, \varphi)$ makes them orthogonal: as a consequence, the spherical harmonics are the natural means for general representation of a function over a spherical surface, analogous to Fourier series for a function in a rectilinear space.

Thus, a gravity field model is made up of the knowledge of GM_E , of a given value for R_0 , and of the values for $C_{n,m}$ and $S_{n,m}$. These coefficients permit to weight the linear combination of each basic spherical function, to get finally a complete modelling of the gravity field. In practice, the field is developed up to a upper degree n_{max} , making the infinite sum over n a finite one. On an intuitive point of view, for a given degree n , the quantity $\frac{20000}{n}$ represents, in kilometers, the spatial scale, or *wavelength*, of the deformation of the continuous body, compared to a homogeneous sphere. For example, deformations of degree 2 have a typical spatial scale of 10000 km: the coefficient of degree 2 and degree 0, $C_{2,0} = -J_2$, corresponds to the dynamical flattening of the central body.

Therefore, $\frac{20000}{n_{max}}$ represents the lowest spatial scale of a deformation which can be described in a gravity field model developed up to the degree n_{max} : a global gravity field model is characterized on the one hand by long and medium wavelengths (500 to 10,000 km), determined in an effective way by space techniques (see hereafter) and linked to global geophysics (Earth's interior), and on the other hand by short and very short wavelengths (50 to 500 km), rather linked to the lithospheric structure.

4 Global Geodynamics

The gravity field is thus described by coefficients $C_{n,m}$ and $S_{n,m}$ which cannot permit to deduce the density at the surface and inside the Earth at any point. Nevertheless, they can enlighten, in a very effective way, many geodynamical and geophysical phenomena which modify the structure of our planet, since these phenomena cause mass transfers with specific spatial wavelengths.

As mentioned in [12], *"The Earth is a dynamic system: it has a fluid, mobile atmosphere and oceans, a continually changing global distribution of ice, snow, and ground water, a fluid core that is undergoing hydromagnetic motion,"* as well as a solid inner core inside, *"a mantle both thermally convecting and rebounding from the glacial loading of the last ice age, and mobile tectonics plates. These processes modify the distribution of mass within the Earth, and consequently affect the gravitational field over a wide range of scales in space and time."*

The relevance of the description (16) is reinforced by the fact that time variations can be taken into account: each coefficient is time-dependent, made

up of a static part and of time variations which raise from hours (atmospheric phenomena) to thousands of years (post glacial rebound). Traditionally, the time-dependency is denoted:

$$C_{n,m}(t) = C_{n,m}(t_0) + \Delta C_{n,m}(t) \quad S_{n,m}(t) = S_{n,m}(t_0) + \Delta S_{n,m}(t) \quad (17)$$

where $C_{n,m}(t_0)$ and $S_{n,m}(t_0)$ are the coefficients of the static gravity field given for an epoch t_0 , and where $\Delta C_{n,m}$ $\Delta S_{n,m}$ contain the time variations of the gravity field, linked with atmospheric pressure, ocean circulation, tides, soil moisture, snow cover...

For example, the time variation of the first zonal parameters J_n can be modelled such as:

$$J_n(t) = J_n(t_0) + \dot{J}_n(t - t_0) + J_n^A + J_n^{SA} + \dots$$

$J_n(t_0)$ characterizes the static gravity field defined at t_0 , $\dot{J}_n$ the secular variation. J_n^A and J_n^{SA} denote the annual and semi-annual variations, respectively, linked to tidal or non tidal phenomena [29]. Other periods should also have to be added.

In particular, the time variations of J_2 are therefore linked to global mass transfers which modify the tensor of inertia of the Earth. [34] have enlightened an extraneous acceleration in the ascending node of [LAGEOS-1 satellite orbit]. This is due to a secular variation of the zonal harmonics J_2 , due partly to the post glacial rebound (see also [28, 31]).

The mass of the Earth, in fact the parameter $\mu = GM_E$, is included in the degree 0 of the geopotential, and is the first parameter constituting a gravity field model, that is $C_{0,0}$. The most effective way to compute it consists in managing the semi-major axis of geodetic satellites, since one can deduce from the third law of Kepler: $3\frac{da}{a} = \frac{d(GM_E)}{GM_E}$. The Lunar Laser Ranging technique (LLR), measuring a distance between the Moon and fixed positions on the Earth, can provide such a determination, even if the Satellite Laser Ranging technique (SLR) provides better results [33], in particular with the LAGEOS-1 [geodetic satellite].

Another example lies in terms of degree 1 ($C_{1,0}$, $C_{1,1}$ and $S_{1,1}$), which give some precious information about the position of the effective geocentre, and on its motion. Orbits of artificial satellites are computed in a frame where the origin is the instantaneous centre of mass. In that frame, the degree one harmonics are by definition equal to zero. The absence of first degree harmonics from the gravity field is equivalent to assuming the geometrical centre of the Earth, to which station positions refer, coincides with the dynamical centre of mass, to which the orbit refers.

However, the centre of figure of the Earth does not, in general, coincide with the geocentre. Adjusting orbits of artificial satellites, on the basis of tracking data which refer to the centre of figure, leads to coefficients of degree 1 not equal to zero. Tracking data define the centre of mass C of the Earth, through its coordinates (x_G, y_G, z_G):

$$\begin{cases}
x_G = R_0 C_{1,1} = \frac{1}{M_E} \iiint_{\mathcal{V}} x_P dM = \frac{1}{M_E} \iiint_{\mathcal{V}} r_P P_{1,1}(\sin \varphi_P) \cos \lambda_P dM \\
y_G = R_0 S_{1,1} = \frac{1}{M_E} \iiint_{\mathcal{V}} y_P dM = \frac{1}{M_E} \iiint_{\mathcal{V}} r_P P_{1,1}(\sin \varphi_P) \sin \lambda_P dM \\
z_G = R_0 C_{1,0} = \frac{1}{M_E} \iiint_{\mathcal{V}} z_P dM = \frac{1}{M_E} \iiint_{\mathcal{V}} r_P P_{1,0}(\sin \varphi_P) dM
\end{cases} \quad (18)$$

The geocentre motion is a consequence of the origin of the inertial frame for orbital computations being taken to be the instantaneous centre of mass of the Earth-atmosphere system [24]. Variations in the geocentre position are a response to mass redistribution of the solid earth, oceans, atmosphere and hydrosphere.

5 Orbital Dynamics

The principle of global measurements of the Earth's gravity field, which has to be realized with appropriate spatial and temporal sampling, consists in analyzing satellite orbit perturbations from a nominal model, thanks to a set of tracking observations: it is an “inverse problem”. This technique of analysis is based on the numerical integration of the equations of motion of the satellite, which include the integration of the variational equations (see hereafter)

Referring to [3], “*Orbit determination is not a well-known task in satellite geodesy. This is mainly due to the fact that orbit determination is often imbedded in a much more general parameter estimation problem, where other parameter types (station positions, Earth rotation, atmosphere...) have to be determined, as well.*”.

5.1 Integrate the Equations of Motion

It is important to note that potentials such as (5) or (6) cause on S a non central force. The keplerian orbital elements of a satellite vary therefore with time, and verify the Lagrange equations, assuming that the disturbing force derives from a potential denoted U , such as the gravity potential (16).

If $\mathbf{E}$ denotes a set of elements which can locate the satellite in space (such as keplerian elements, but also such as positions and velocities), these equations read, if initial values $\mathbf{E}_0$ are given at an initial epoch t_0 :

$$\boxed{
\begin{aligned}
\frac{d\mathbf{E}}{dt} &= f(\mathbf{E}, (\sigma_k)_{k=1,n}) \\
\mathbf{E}(t_0) &= \mathbf{E}_0
\end{aligned}
} \quad (19)$$

In these equations, the parameters (σ_k) describe the environment of the satellite, and in particular the gravity field through the coefficients $C_{n,m}$ and $S_{n,m}$.

The orbit of a satellite can therefore be generated (extrapolated), thanks to numerical integration algorithms, and effects of given parameters can be enlightened (Fig. 4). The secular perturbations are the dominant perturbations of geodetically useful satellites - that is, those high enough not to suffer excessive atmospheric drag, but low enough to be perceptibly perturbed by the complex structure of the Earth's gravity field.

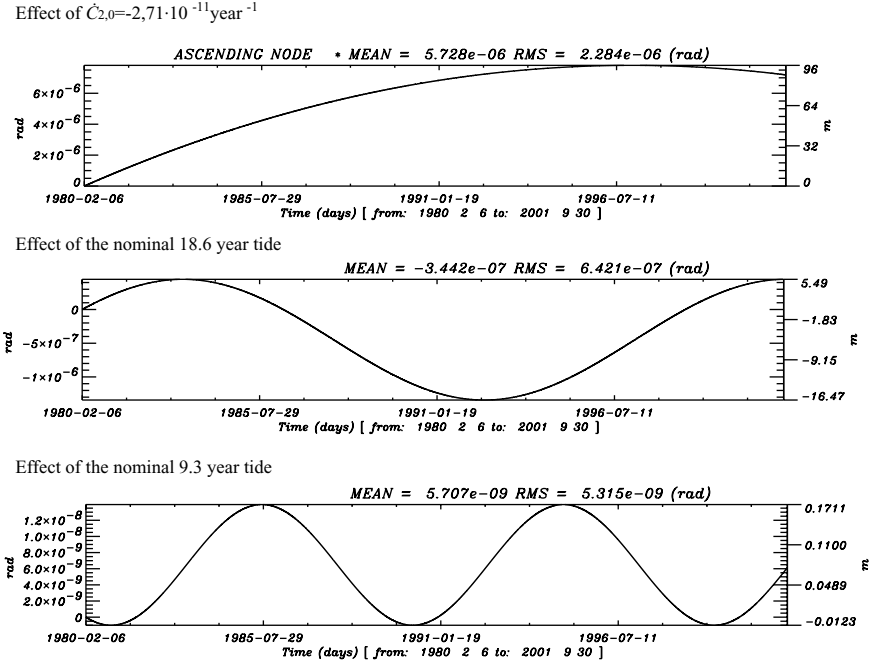


Fig. 4. Effects of geodynamical parameters on the ascending node of [LAGEOS-1 satellite orbit]. Extracted from [10]. In this figure are shown, for three given parameters, the long period evolutions of the ascending node over a period of 20 years. Each curve is the difference of two extrapolations: the first curve shows the impact of the secular variation of $C_{2,0}$, the second of the 18.6 year equilibrium tide [29] which is equivalent to 1.23 cm of water, the third of the 9.3 year tide which is equivalent to 0.02 cm of water

Many periods are involved during the numerical integration of the equations of motion, and mainly the periods of revolution of the satellites and of the Earth, and their multiples. However, the very small orbital signatures of mass redistribution within the Earth are the result of a long integration process that gradually changes the variables characterizing the satellite motion. A

technique of averaging the equations of motion provides very long orbital arcs, that is of thousands of days and more (compared to several days in classical approaches). It has been revealed as the most effective method for determining and decorrelating secular and long period variations in the Earth's gravity field.

5.2 Computing from Space the Coefficients of the Gravity Field

We present now the general method which permits to compute the coefficients of the gravity field from the perturbations observed on an orbit.

Adjusting an orbit on given tracking observations consists in finding the best parameters $\Sigma = (\sigma_k (k = 1, p))$ -“best” in a sense to be defined- which enable to build the closest orbit with respect to these n observations $O(t)$ sampled in time. Different kinds of observations are available, whatever the used technique: radial positions for the SLR technique, 3 spatial coordinates for GPS (Global Positioning System), radial velocities for the DORIS system [35].

The integration of the equations of motion permits to deduce, from $\mathbf{E}(t)$, quantities O_{th} directly comparable to the observations $O(t)$. For the SLR technique, $O(t)$ is the distance between the position of a satellite and a tracking station on the Earth, and O_{th} is the same distance computed thanks to mathematical models.

If the modelling were perfect, one would get, at any epoch t_i when an observation is available, $O(t_i) = O_{th}(t_i)$. Since it is not the case, the function to be minimized is $O(t_i) - O_{th}(t_i)$, seen as a function of the parameters Σ . In the well-known least-square method, the useful estimation method in space geodesy [32], an a priori solution Σ_0 of Σ is required, as well as the following matrix:

$$A = \begin{pmatrix} \frac{\partial O_{th}}{\partial \sigma_1}(\Sigma_0, t_1) & \cdots & \frac{\partial O_{th}}{\partial \sigma_p}(\Sigma_0, t_1) \\ \vdots & \cdots & \vdots \\ \frac{\partial O_{th}}{\partial \sigma_1}(\Sigma_0, t_i) & \cdots & \frac{\partial O_{th}}{\partial \sigma_p}(\Sigma_0, t_i) \\ \vdots & \cdots & \vdots \\ \frac{\partial O_{th}}{\partial \sigma_1}(\Sigma_0, t_n) & \cdots & \frac{\partial O_{th}}{\partial \sigma_p}(\Sigma_0, t_n) \end{pmatrix} \quad (20)$$

To build the A matrix, the quantities $\frac{\partial O_{th}}{\partial \sigma_k} = \frac{\partial O_{th}}{\partial \mathbf{E}} \frac{\partial \mathbf{E}}{\partial \sigma_k}$ are required. If $\frac{\partial O_{th}}{\partial \mathbf{E}}$ can be obtained in an explicit way (since, for example, positions can be expressed with respect to keplerian elements), and can be seen as sensibilities of tracking measurements to orbital parameters, the solely available quantities for the second factor are $\frac{\partial}{\partial \sigma_k} \left(\frac{d\mathbf{E}}{dt} \right) = \frac{\partial f}{\partial \mathbf{E}} \frac{\partial \mathbf{E}}{\partial \sigma_k} + \frac{\partial f}{\partial \sigma_k}$, referring to (19). Assuming classically, even if it can be questioned from a purely mathematical point of view, that $\frac{\partial}{\partial \epsilon_k} \left(\frac{d\mathbf{E}}{dt} \right) = \frac{d}{dt} \left(\frac{\partial \mathbf{E}}{\partial \epsilon_k} \right)$, one gets differential equations

which can be integrated at the same time as the equations of motion. These equations are called *variational equations*.

The solution is

$$\Sigma = \Sigma_0 + (A^T P A)^{-1} A^T P (\mathbf{E}(t_i) - f(\Sigma_0, t_i)) \quad (21)$$

where P can be defined as the identity matrix, except if an error budget exists for the observations. Several iterations are often necessary to get a stable solution.

The temporal variations of the geopotential produce orbital signals that are of the order of several mas/yr on the node and eccentricity, notably (Figs. 5 and 6, respectively). As an example of determining global geodynamical parameters from orbit perturbations, Fig. 5 shows the time variations of effective values of $J_2 = -C_{2,0}$ and $J_3 = -C_{3,0}$ obtained from the analysis of 2-year orbital arcs of several geodetic satellites [9]. The mean tendency for J_2 is a

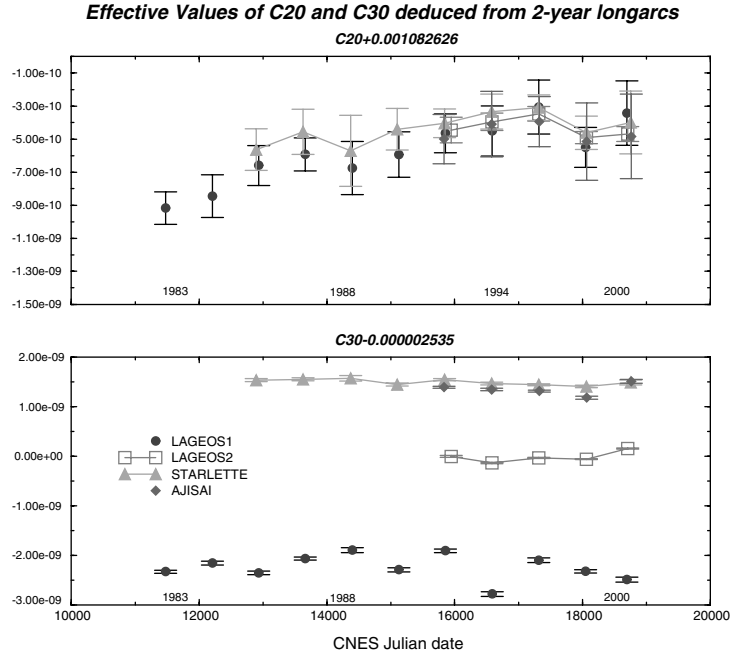


Fig. 5. Adjustment of effective values of $J_2 = -C_{2,0}$ and $J_3 = -C_{3,0}$ on the orbits of several satellites. For a given satellite, an effective value of J_2 corresponds in fact to the influence of a linear combination of corrections to all even zonal parameters and other parameters of degree 2, such as tidal parameters; an effective value of J_3 corresponds to the influence of a linear combination of corrections to all odd zonal parameters and other parameters of degree 3 [14]

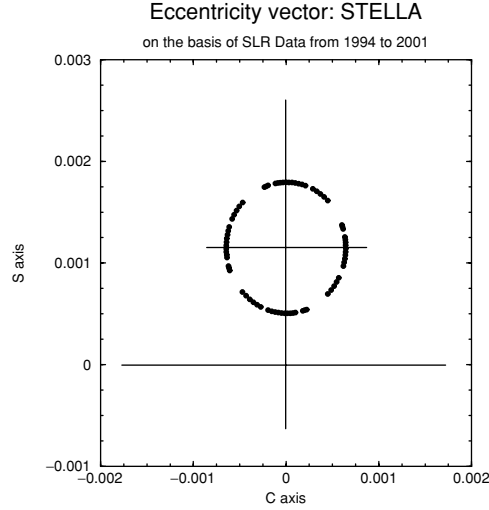


Fig. 6. Evolution of the eccentricity vector of the STELLA satellite. Adapted from [11]. The eccentricity vector is defined as ($C = e \cos \omega$, $S = e \sin \omega$). The rectangular coordinates of the centre of the circle described by the eccentricity vector are proportional to $\frac{J_3}{J_2}$

straight line, corresponding to the effect of the post-glacial rebound. The origin of the rupture of this tendency, around 1987 and 1997, is anything but bright. It could be due to a redistribution, well located in time, of oceanic mass [7,13] and [8]. Studies have to be continued, in particular to establish to what extent there is, or not, a strong correlation with El Niño events ([15] is a good introduction to the physical mechanisms involved in these events).

6 Current Purposes

Analysis of the orbits of the usual geodetic satellites, such as LAGEOS-1 or STARLETTE, have already yielded important information concerning the gravity field of the Earth. These satellites, with low area to mass ratio to minimize non gravitational accelerations, are effectively passive satellites dedicated to studies of geodesy and geodynamics. Temporal variations have been enlightened in many studies (for example [18] or [36]). But, in addition to long wavelengths (1,000 to 10,000 km), a dramatic improvement in the description of the temporal variations of the gravity field, including the shortest wavelengths (200 to 1,000 km), has started recently with new space missions.

6.1 Combined Gravity Field Models

As a consequence of accumulating satellite observations and corresponding normal matrices ($A^T P A$), a great accuracy has been reached for gravity

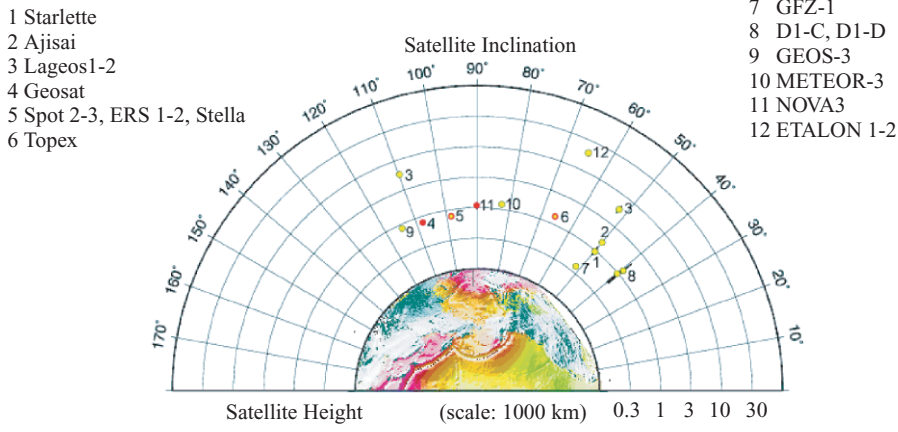


Fig. 7. Satellites and orbits employed for GRIM5-S1 global gravity field recovery

coefficients of low degrees (typically up to degree 10). For example, the coefficient J_2 ($1.082 \cdot 10^{-3}$) has been determined with 8 digits corresponding to an accuracy of about 10^{-11} . Let us recall that global mass transfer (ocean, atmosphere and solid earth) have signatures of 10^{-10} at degree 2, and less.

For example, the GRIM5-S1 model [4] is based on a classical approach, the analysis of gravitational satellite orbit dynamic (Fig. 7). Spatial resolution of that satellite-only model (in opposition to other which can include gravimetric and altimetric data) is of about 1200 km at the Earth's surface. Table 1 gives the first values of the GRIM5-S1 model, to be compared, for example, with the first values of the gravity field of Mars.

To improve the accuracy of global gravity field models at short wavelengths (lower than 500km), measurements obtained from gravimetric data (mainly over lands, and partially over the oceans) and from altimetric data (direct measurement from space of ocean topography), have been processed in a way to supply additional information to the existing purely satellite normal equations ($A^T P A$). The GRIM5-C1 model is an example of that combined models [19].

Moreover, oceanographic satellite missions, such as ERS lead by the European Space Agency (ESA) in the 1990s, provided extremely fine information about the structure of the oceanic lithosphere thanks to the corresponding signature they breed on the mean sea surface topography (MSST) [6]. Thanks to a series of altimetric missions over about 20 years, from GEOSAT to ENVISAT, the quality of the knowledge of the geoid, at medium and short wavelengths (400-800 km) has become much better over the oceans than over lands, particularly at high latitudes (by a factor from 2 to 3).

The german CHAMP mission, launched on July 15, 2000, has been designed to fill that gap over lands: the choices of its orbits (inclination of 87° , altitude of 450 km), of the tracking systems (GPS and SLR), and of on-board

Table 1. Comparisons of first terms characterizing the gravity fields of the Earth and Mars

	The Earth (GRIM5)	Mars (GMM2B)
$GM \text{ (} m^3.s^{-2} \text{)}$	0.39860044144982E+15	0.42828371901284E+14
$R_0 \text{ (km)}$	6378.13646	3397
$\frac{1}{f}$	0.29825765E+03	0.19113720E+03
$C_{2,0}$	-0.48416511557015E-03	-.87450547081842E-03
$C_{3,0}$	0.95857491647196E-06	-.11886910646016E-04
$C_{4,0}$	0.53978784172512E-06	0.51257987175466E-05

equipments (3-axe accelerometer) insure an impressive success in the determination of homogeneous (including lands and polar areas) and precise models (several versions) at short wavelengths (400km).

At this altitude, the non gravitational forces, such as the atmospheric drag, can not be modelled with a great accuracy, at any place and time. That is why accelerometers are built-in. Accelerometers, built by ONERA (Office National d'Etudes et de Recherches Aérospatiales, Paris), provide information for separating non-conservative accelerations from gravity and help to prevent aliasing into the adjusted coefficients characterizing the static gravity field. Moreover, the on-board GPS receivers provide a permanent 3-D positioning, to be compared with the SLR measurements which permits a very accurate tracking but only on the radial direction.

First results and complete gravity field models including data obtained from the CHAMP mission start to come [25, 30].

6.2 The New Missions GRACE and GOCE

If the GPS CHAMP receivers, coming from JPL (Jet Propulsion Laboratory, Pasadena, Ca.), make that mission be a SST-hl mission (Satellite to Satellite Tracking - high low), even more precise measurements are available when the mission is a SST-ll one (Satellite to Satellite Tracking - low low): it is the case of the recent american-german GRACE mission, the two GRACE satellites having been launched at an altitude of 450 km on March 17, 2002, and scheduled for 5 years. The distance between the twin satellites is measured by a micrometer-precision ranging system, in addition to GPS, SLR and accelerometers. Variations of that distance form the basis of a new contribution in the determination of the gravity field. GRACE is rather dedicated to the time variabilities of low and high degree gravity coefficients, making that mission the first tool for quantifying and modelling global mass transfers.

The next step in gravity missions consists in measuring, in addition to U (with CHAMP) and to the partial derivatives of U (with GRACE), the second derivatives of U . In that goal, the GOCE mission, planned for 2007 by ESA at an altitude of 270km, will provide an unprecedented accuracy for the high

Table 2. Precisions required in space geodesy, in terms of gravity ($1\text{mgal} = 10^{-5}\text{m.s}^{-2}$) and geoid heights (in centimeters) from the reference ellipsoid

Accuracy on:	Geoid	Gravity	Spatial Resolution
Oceanic circulation			
High spatial scales	2 cm		60–250 km
Low spatial scales	< 1 cm		1000 km
Geodynamics			
Continental lithosphere		1–2 mgals	50–400 km
Mantle composition		1–2 mgals	100–5000 km
Oceanic lithosphere		5–10 mgals	100–200 km
Upheaval processes		5–10 mgals	
Geodesy			
Vertical motions	2 cm	1–5 mgals	100–200 km
Lower structures of polar caps		1–5 mgals	50–100 km
Global Height	< 5 cm	1–5 mgals	50–100 km

degree and order gravity coefficients. Such SGG measurements (for “Satellite Gravity Gradient”), will give access from space to data only available from ground up to now. As a consequence, this will homogenize the description of the short wavelengths of the geoid, and in particular over the oceans. This will make GOCE fundamental for better exploring the ocean circulation at a global level. Table 2 describes different resolutions and precisions that are still in question to improve geophysical and oceanographical models.

6.3 Towards an Alternative to Spherical Harmonics for Short Spatial Wavelengths

The success of the spherical harmonics in space geodesy, reinforced since a few decades, comes from the capability of modern softwares to take into account a very high number of coefficients, and to estimate them, on the basis on variational equations. Nevertheless, even if the description of the gravity field is more and more precise, as the spatial resolution of the models decreases, the physical meaning of a given coefficient may disappear: with spherical harmonics, the effect of a gravity anomaly (due to a mountain for example) is involved in many coefficients, boiling down to nothingness an interpretation of a given coefficient from a geophysical point of view. Current studies aim at using other modellings of the gravity field, such as wavelet decompositions [27], which can locate in a very precise way gravity anomalies.

Let us note, finally, that other modellings, derived from spherical harmonics, can be used when the central body is not a sphere even in a first approximation. For example, ellipsoidal harmonics have been used to study the gravity field of the Eros asteroid [17].

7 Conclusion

With only a few exceptions on the Earth's surface, the history and spatial pattern of mass transfer over large spatial scales are often not accessible to direct observations. Space geodetic techniques, however, have the capability of monitoring certain direct and global consequences of the mass transport, including Earth's variations in rotation, gravity field, and in its geocentre motion: geodesy is a multidisciplinary science [5]. As a consequence, relationships between astronomers and geophysicists are now essential, and have to be developed.

The GRACE mission is currently measuring the temporal variations of the gravity field. The GOCE mission will enhance our knowledge of the global static gravity field and of the geoid by orders of magnitude. With these new data, in addition to those obtained from the classic tracking of geodetic satellites with the GPS constellation and the SLR network, a whole range of fascinating new possibilities is now opened for solid earth physics, oceanography and sea-level research: the study of our alive planet from space is now so precise that the physical meaning of each harmonic gravity coefficient has to be questioned; that is why theoretical aspects of the study of a gravity field have to be kept in mind.

Finally, let us note that the concepts developed for the Earth have also be used to study the gravity field of other planets, thanks to perturbative methods, provided that spacecrafts have flown over these planets. For Venus, a 180th degree and order model has been built, on the basis of the Magellan data [1]. The high resolution comes from highly eccentric orbits, and, as a consequence, from low altitude periapses. For the Moon, a recent model can be found in [21], waiting for the results of the SELENE mission [23]. For Mars, the most recent model has been deduced from Mars Global Surveyor [22], and should be improved thanks to data obtained with Mars Express. Studying the gravity field of a planet is effectively a task belonging to astronomy and geodesy.

Acknowledgements

The authors thank the 'Centre National d'Etudes Spatiales' for financial support of Florent Deleflie. We would like to acknowledge Jean Souchay, the scientific organizer of the School entitled *Dynamique des corps célestes non ponctuels et des anneaux*, for inviting us to give this lecture. We would also like to thank Gilles Métris for providing helpful suggestions, and François Barlier for helpful discussions and permanent encouragement.

References

1. Barriot J.-P., Valès N., Balmino G., Rosenblatt P., *A 180th degree and order model of the Venus gravity field from Magellan line of sight residual Doppler data*, Geophysical Research Letters, 25 (19), 3743-3746, 1998
2. Beutler G., M.R. Drinkwater, R. Rummel, Rudolf von Steiger (eds.), *Earth Gravity Field from Space - from Sensors to Earth Sciences*, ISBN 1-4020-1408-2, Kluwer Academic Publishers, 2003
3. Beutler G., T. Schildknecht, U. Hugentobler, W. Gurtner, *Orbit Determination in Satellite Geodesy*, Adv. Space Res., 31, 8, pp. 1853-1868, 2003
4. Biancale R., G. Balmino, J.-M. Lemoine, J.-C. Marty, B. Moynot, F. Barlier, P. Exertier, O. Laurain, P. Gegout, *A new global Earth's gravity field model from satellite orbit perturbations: GRIM5-S1*, Geophys. Res. Letters, 27, 22, 3611-3614, 2000
5. Cazenave A., K. Feigl, *Formes et mouvements de la Terre. Satellites et géodésie*, CNRS Editions, 1994
6. Cazenave A., *The Geoid and Oceanic Lithosphere*, in 'Geoid and its Geophysical Interpretations', P. Vanicêk Ed., CRC Press Inc., 1994
7. Cazenave A., R.S. Nerem, *Redistributing Earth's Mass*, Science, 297, 783-784, 2002
8. Cox C., B.F. Chao, *Detection of a Large-Scale Mass Redistribution in the Terrestrial System Since 1998*, Science, 297, 831-833, 2002
9. Deleffie F., *Théorie semi-analytique des mouvements quasi-circulaires moyens en mécanique spatiale - Applications aux satellites géodésiques*, Thèse de doctorat de l'Observatoire de Paris, 2002
10. Deleffie F., P. Exertier, G. Métris, P. Berio, O. Laurain, J.-M. Lemoine, R. Biancale, *Why the 18.6 year tide cannot explain the change of sign observed in $J_2\dot{\theta}$?*, Advances in Geosciences, 1, 103-108, 2003
11. Deleffie F., G. Métris, P. Exertier, *Long period variations of the eccentricity vector for near circular orbits around a non spherical body*, submitted to Celest. Mech., 24p, 2003
12. Dickey J.O., *Time Variable Gravity: An Emerging Frontier in Interdisciplinary Geodesy*, IAG Symposia, 123, Sideris (ed.), pp 1-3, Springer-Verlag, 2001
13. Dickey J. O., S. L. Marcus, O. de Viron, and I. Fukumori, *Recent Earth oblateness variations: Unraveling climate and postglacial rebound effects*, Science, 298, 1975-1977, 2002
14. Eanes R.J., S. Bettadpur, in *Global Gravity Field and Its Temporal Variations*, pp. 30-41, Springer-Verlag, New York, 1996
15. Enfield D.B., *El Niño, past and present*, Reviews of Geophysics, 27, 1, 158-187, 1989
16. Exertier P., *Geopotential From Space Techniques*, Celest. Mech., 57, 137-153, 1993
17. Garmier R., J.-P. Barriot, *Ellipsoidal harmonic expansion of the gravitational potential: theory and application*, Celest. Mech., 79, 4, 2001
18. Gegout P., and A. Cazenave, *Temporal variations of the Earth gravity field for 1985-1989 derived from Lageos*, Geophys. J. Int., 114, 347-359, 1993
19. Gruber T., A. Bode, C. Reigber, P. Schwintzer, G. Balmino, R. Biancale, J.-M. Lemoine, *GRIM5-C1: Combination solution of the global gravity field to degree and order 120*, Geophys. R. Let., 27, 24, 4005-4008, 2000

20. Kaula W.M., *Theory of satellite geodesy*, Blaisdell Publishing Company, Waltham, Massachusetts, pp 8–10, 1966
21. Konopliv A. S., Asmar S. W. Carranza E., Sjogren W. L., and Yuan D. N., *Recent Gravity Models as a Result of the Lunar Prospector*, *Icarus*, 150, 1–18, 2001
22. Lemoine F.G., D. E. Smith, D. D. Rowlands, M. T. Zuber, G. A. Neumann, and D. S. Chinn, *An improved solution of the gravity field of Mars (GMM-2B) from Mars Global Surveyor*, *J. Geophys. Res.*, 106(E10), 23359–23376, 2001
23. Matsumoto K., Heki K., and Hanada H., *Global Lunar Gravity Field Recovery from SELENE*, *Proc. International VLBI Service for Geodesy and Astrometry 2002 General Meeting*, 2002
24. Moore P., J. Wang, *Geocentre Variation from Laser Tracking of LAGEOS1/2 and Loading Data*, *Adv. Space Res.*, 31, 8, pp. 1927–1933, 2003
25. Moore P., J.F. Turner, Z. Qiang, *CHAMP Orbit Determination and Gravity Field Recovery*, *Adv. Space Res.*, 31, 8, pp. 1897–1903, 2003
26. Morando B., *Mouvement d'un satellite artificiel de la Terre*, Gordon & Breach, 1974
27. Panet I., M. Diamant, O. Jamet, *A wavelet based representation of the gravity field*, 2nd International GOCE user workshop, ESA, March 2004
28. Peltier W.R., *Constraint on Deep Mantle Viscosity from Lageos acceleration data*, *Nature*, 304, 434–436, 1983
29. Ray R.D., and D.E. Cartwright, *Satellite altimeter observations of the Mf and Mm ocean tides, with simultaneous orbit corrections*, in *Gravimetry and Space Techniques Applied to Geodynamics and Ocean Dynamics*, edited by B.E. Schutz, A. Anderson, C. Froidevaux, and M. Parke, 69–78, American Geophysical Union Geophysical Monograph 82, IUGG Vol. 17, Washington D.C., 1994
30. Reigber C., P. Schwintzer, K.-H. Neumayer, F. Barthelmes, R. König, C. Förste, G. Balmino, R. Biancale, J.-M. Lemoine, S. Loyer, S. Bruinsma, F. Perosanz, T. Fayard, *The CHAMP-only Earth Gravity Field Model EIGEN-2*, *Adv. Space Res.*, 31, 8, pp. 1883–1888, 2003
31. Rubincam D.P., *Post-glacial rebound observed from Lageos and the effective viscosity of the lower mantle*, *J. Geophys. Res.*, 89, 1077–1087, 1984
32. Sillard P., *Estimation par moindres carrés*, Lavoisier Editions, 2001
33. Smith D.E., R. Kolenkiewicz, P.J. Dunn, M.H. Torrence, *Earth scale below a part per billion from Satellite Laser Ranging*, *IAG Symposia*, 121, Schwarz (ed.), pp 3–12, Springer-Verlag, 2000
34. Yoder C.F., J.G. Williams, J.D. Dickey, B.E. Schutz, R.J. Eanes, B.D. Tapley, *Secular variation of Earth's gravitational harmonics J2 coefficient from LAGEOS and non tidal acceleration of Earth rotation*, *Nature*, 303, 757–762, 1983
35. Tavernier G., P. Willis, *The International DORIS Service*, *Geodesist's Handbook*, Springer-Verlag, 2004.
36. Zhu Y., C.K. Shum, M.K. Cheng, B.D. Tapley, and B.F. Chao, *Long-period variations in gravity field caused by mantle anelasticity*, *J. Geophys. Res.*, 101, B.5, 11,243–11,248, 1996

Asteroids from Observations to Models

D. Hestroffer¹ and P. Tanga²

¹ IMCCE, UMR CNRS 8028, Observatoire de Paris. 77 Av. Denfert Rochereau
F-75014 Paris, France
hestro@imcce.fr

² Observatoire de Nice, Dépt. Cassini, B.P. 4229, F-06304 Nice, France
tanga@obs-nice.fr

Abstract. We will discuss some specific applications to the rotation state and the shapes of moderately large asteroids, and techniques of observations putting some emphasis on the HST/FGS instrument.

1 Introduction

Although their name suggest they are point-like, asteroids are from long time well known to show variations in their lightcurves with shape and rotation [1]. Observed lightcurves can for instance be explained by spinning tri-axial ellipsoids, but even better by convex shapes in uniform rotation [2,3]. Also the rotation period of these bodies seem to have some connection with their size. For instance asteroids larger than approximately 0.15 km do not spin faster than ≈ 10 cycles/day (see Fig. 4). A. Harris [4] has suggested that this limit is not the tail of some statistical distribution but does correspond to the limit of disruption of a gravitationally bound and cohesionless *rubble pile*. We will discuss in the following on the inversion of asteroids lightcurves, on rotational state of asteroids in general, on possible figures of equilibrium for rubble pile asteroids, and illustrate how we can derive information on asteroids shape and size from high resolution observation with the HST/FGS interferometer.

2 Lightcurves

One of the first and most important source of knowledge on the asteroids physical properties was obtained from photometric lightcurve observations. These reveal the asteroid's rotation period, their brightness variation due to their non spherical shape, albedo spots, light scattering of their surface, ... Two typical examples of asteroids lightcurve are given in Fig. 1. In particular one can see from Sylvia's composite lightcurve that the brightness variation is

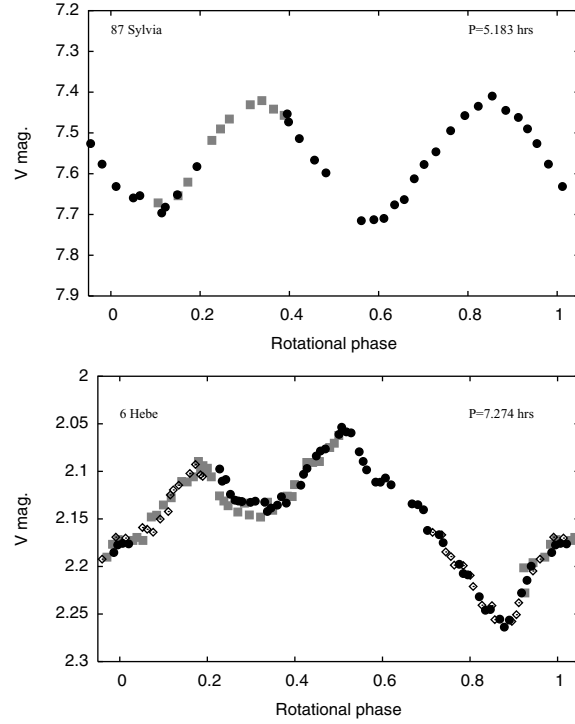


Fig. 1. Composite lightcurves of two asteroids from the Asteroid Photometric Catalogue [2]. *Upper panel:* (87) Sylvia observed on 1987 Feb. 3.3 (*filled circles*), and Feb. 6.3 (*filled squares*), by Weidenschilling et al. (1990). *Lower panel:* (6) Hebe observed on 1987 Jun. 18 (*filled squares*), Jun. 23 (*filled circles*), and Jun. 27 (*open squares*) by Hutton & Blain (1988)

periodic, and that there are two – almost identical – maxima and minima. Thus Sylvia’s brightness variation can be well be approached by a tri-axial ellipsoid in rotation. So that analysis of several lightcurves obtained at different apparitions provide the pole orientation and the axis ratio of the ellipsoid shape model. In the more general case however, lightcurves are not always so smoothly sinusoidal, but more irregular and asymmetric with additional extrema, ... (see the lightcurve of Hebe in Fig. 1 for an illustration). Promising results in the lightcurve inversion problem have been obtained recently [5].

3 Rotation

Considering a freely rotating rigid body (Euler’s spinning top), integration of the Euler equations yields the orientation in the frame of the body of the instantaneous velocity $\mathbf{\Omega}$, and the orientation – via the Euler angles (ϕ, θ, ψ) –

of the body in a inertial frame [6]. As noted in [7] such a dynamical system has two – generally non commensurable – frequencies, so that the body never shows the same aspect in time. Although this exists in the solar system, there are however only a very few small bodies that are known to clearly show (or suspected to show) such a spin state [8]: among them one comet P/Halley [9], and one asteroid, asteroid (4179) Toutatis [10]. Moreover all these objects have relatively long rotation (spin) period. We shall see later the reason for this lack of (fast) precessing bodies.

Let us remind that integration of the Euler equations:

$$\begin{aligned} I_1 \dot{\omega}_1 + (I_3 - I_2) \omega_2 \omega_3 &= 0 \\ I_2 \dot{\omega}_2 + (I_1 - I_3) \omega_3 \omega_1 &= 0 \\ I_3 \dot{\omega}_3 + (I_2 - I_1) \omega_1 \omega_2 &= 0 \end{aligned} \quad (1)$$

and

$$\begin{aligned} \dot{\phi} &= \frac{\omega_1 \sin \psi + \omega_2 \cos \psi}{\sin \theta} \\ \dot{\psi} &= \omega_3 - \cos \theta \dot{\phi} = \omega_3 - \frac{\cos \theta (\omega_1 \sin \psi + \omega_2 \cos \psi)}{\sin \theta} \\ \dot{\theta} &= \omega_1 \cos \psi - \omega_2 \sin \psi \end{aligned} \quad (2)$$

gives the orientation of the asteroid principal axes with respect to an inertial frame at any epoch t [6,7]. In a previous chapter by Tokieda we have also seen that the conservation of kinetic energy and angular momentum provides two integral relations:

$$\begin{aligned} 2E &= I_1 \omega_1^2 + I_2 \omega_2^2 + I_3 \omega_3^2 \\ M^2 &= I_1^2 \omega_1^2 + I_2^2 \omega_2^2 + I_3^2 \omega_3^2 \end{aligned} \quad (3)$$

The (x_1, x_2, x_3) body-frame (see Fig. 2) is a right-handed frame associated to (x_s, x_i, x_l) , and with this choice of indexing we have put $I_1 > I_2 > I_3$. In general the Euler angles are given with respect to a inertial frame where the z -axis is – for commodity – aligned with the angular moment, and are understood as rotation, precession and nutation angles. In that case both angles θ and ψ are periodic functions of commensurable period. Putting:

$$k^2 = \frac{(I_2 - I_1)(2E I_3 - M^2)}{(I_3 - I_2)(M^2 - 2E I_1)} < 1$$

and making use of the elliptic integral of the first kind $K = \int_0^{\pi/2} (1 - k^2 \sin^2 u)^{-1/2} du$, the period of the rotation angle $\tilde{\psi}$ is [7]:

$$T_{\text{rot}} = 4 K \sqrt{\frac{I_1 I_2 I_3}{(I_3 - I_2)(M^2 - 2 E I_1)}} \quad (4)$$

while the period of nutation θ is $T_{\text{nut}} = T_{\text{rot}}/2$. Last, Landau & Lifchitz [7] have shown that the angle ϕ can be obtained as a sum of two periodic functions $\phi(t) = \phi_1(t) + \phi_2(t)$ where the period of $\phi_1(t)$ is exactly $T_{\text{rot}}/2$ and the period of $\phi_2(t)$ is T' , which in the general case is not commensurable to T_{rot} . The latter period can be obtained from $2\pi T_{\text{rot}} / \int_0^{T_{\text{rot}}} \dot{\phi}(t) dt$. For celestial bodies one may prefer to express the Euler angles in the frame of the ecliptic J2000. In that case the two fundamental frequencies are mixed, and the Euler angles can be expressed as a sum of periodic functions of non-commensurable periods. Interestingly, none of the Euler angles is varying uniformly with time in the general case, which gave the name 'tumbling' to the asteroid Toutatis spin state³. In the case of Toutatis, two of the principal axes have approximately the same moment of inertia ($I_2 \sim I_3$), so that the nutation is small ($\dot{\theta} \sim 0$) and the precession and rotation are circulating with non uniform velocity as shown in Fig. 3. Note also that, considering the readily inequalities $2 E I_3 < M^2 < 2 E I_1$, we can distinguish long-axis mode rotation (LAM, which is the case for Toutatis⁴ when $M^2 > 2 E I_2$, from short-axis mode rotation (SAM) ($M^2 < 2 E I_2$). This denomination reflects the fact that the instantaneous rotation axis is closer to the long axis, or the short axis of the body (excluding "medium-axis rotation" that anyway are not stable).

One can thus expect that rigid bodies that suffer collisions and/or external perturbation can be in complex rotational state (misaligned rotation). In the general case however the asteroids do not show such large free precession/nutation tumble or wobble over time span of typically a few days⁵. In contrast, it is observed that they are spinning around an axis that 1) approximately coincides with the direction of maximal inertia, and 2) which direction is approximately fixed in (inertial) space. Asteroids, as the majority of celestial objects, are not infinitely rigid bodies but are deformed under stresses. When not in principal axis spin-state, there is a loss in kinetic energy: during each wobble period a portion of the stored stress-strain energy is dissipated in the asteroid's interior. Since the angular momentum is conserved and $2T = M^2/I$ is decreasing, this dissipation is balanced by a spin state that evolves asymptotically in time toward a rotation along the axis of largest inertia (I_1 with the notation adopted here). The timescale of this damping process has been analyzed by [11]. It depends on the spin rate Ω , the shape K_1 and size D of the body, its density ρ , and of course on its rigidity μ :

³See an animation on URL

<http://www.star.ucl.ac.uk/apod/solarsys/cap/ast/toutspin.htm>

⁴This is the reason of our choice of ordering $I_1 > I_2 > I_3$, when the most often used one is to put $I_3 > I_2 > I_1$. With the presently adopted convention we only ensure that the rotation is associated to the Euler angle ψ and axis x_3 etc.

⁵Precession due to planetary perturbation is not discussed here.

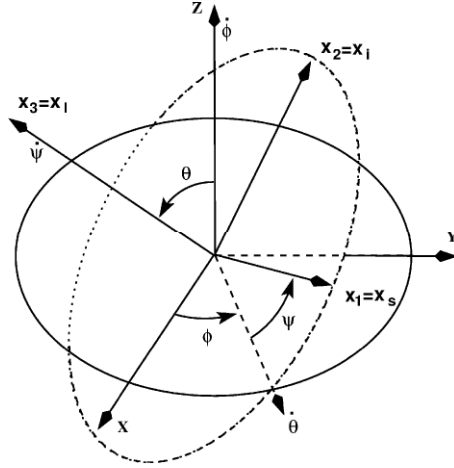


Fig. 2. Euler angles (ϕ, θ, ψ) of the body frame (x_1, x_2, x_3) given in an inertial reference frame (x, y, z)

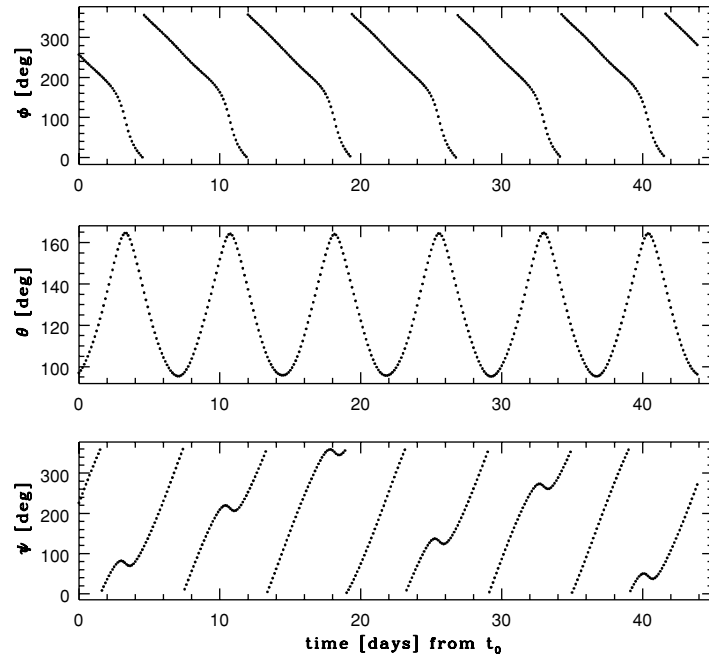


Fig. 3. Euler angles for (4179) Toutatis. Note the non uniformity of the circulation for angles ϕ and ψ

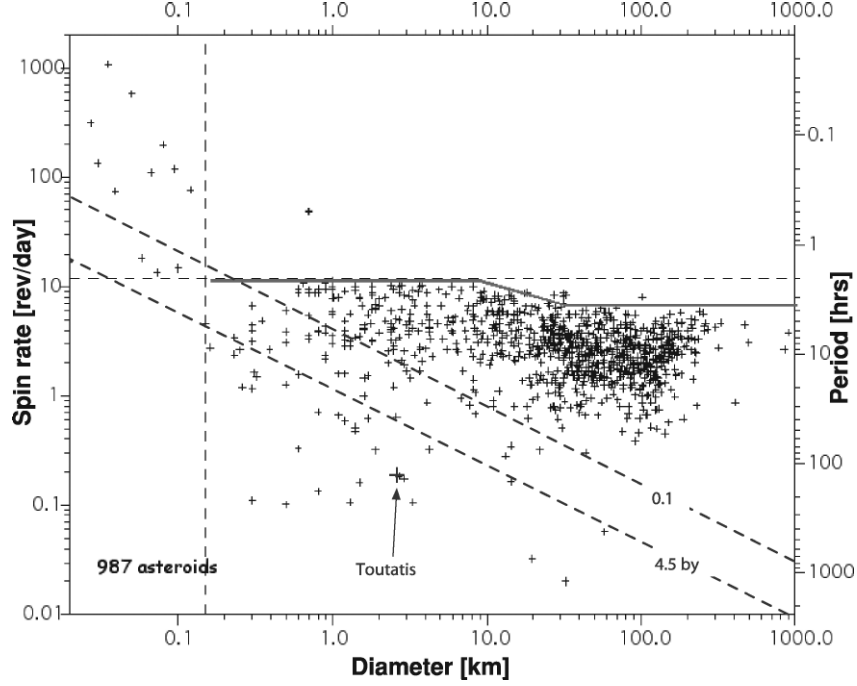


Fig. 4. Spin periods versus size of 987 known asteroids (from P. Pravec.) and damping time scales (*dashed lines*). A limit in rotation rate for large bodies has been drawn by hand

$$\tau_d \sim \mu Q / (\rho K_1^2 (D/2)^2 \Omega^3) \quad (5)$$

where Q is a quality factor expressing the ratio per wobble period of the energy contained in the oscillation to the energy lost. If the body has no rigidity it will instantaneously align its spin axis, and if it is infinitely rigid there is no evolution. Considering values typical of asteroids, [12] alternatively gives the damping timescale by:

$$\tau_d \simeq P^3 / (17^3 D^2) \quad [\text{in } 10^9 \text{ yr}]$$

where P is the rotation period given in hour, and D is the diameter in kilometer. Asteroids densities are (roughly) in the range 1–3, their shape gives $K_1^2 \simeq 0.01 - 0.1$. The value of the product μQ is not known and is still in debate. Burns & Safranov used a value of $\approx 3 \times 10^4$ (cgs units), on another hand Harris [12] – by considering rubble-piles instead of non-porous solid rock and based on available estimates from an analysis of Phobos orbit – adopted a value of 5×10^{12} Pa, while Efroimsky [13] states that the damping time would be 2 to 6 orders of magnitude shorter. Damping timescales with values adopted from Harris are plotted in Fig.4. One notes that either small asteroids and/or long period asteroids are more likely candidates for tumbling rotations – as is

the case for Toutatis – since their damping timescale is close to the age of the Solar System. On the other hand all asteroids larger than ≈ 1 km in diameter and with a rotation period of less than about 10 hours should have their spin axis aligned with the largest moment of inertia axis after 100 million years. This damping hence explains that the vast majority of asteroids have single periodic lightcurves, and exhibit an invariable spin direction aligned with – or close to – their shortest axis of figure and also their axis of angular momentum (We do not mention long period precession/nutation which are not easily accessible with present ground-based photometry).

4 Figures of Equilibrium

As seen previously, we know that (disruptive) catastrophic collisions can result in a *rubble-pile* asteroid, i.e. gravitationally bound re-accumulated aggregates with no internal cohesion [14]. It has been shown in [15] that asteroids could be fractured but still gravitationally bound without losing their integrity from the centrifugal forces. Here we are interested in some particular class of asteroids that excludes the few “giants” bodies (Ceres, Pallas, Vesta,...), and the smallest ones which are supposed to be fully cohesive rocks. The limits in size of the rubble-pile category are not well-defined but could be in the range 1km to 300km [16]. Pravec & Harris [17], noting that fast rotators can hardly have negligible tensile strength, claim that bodies larger than ≈ 0.15 km in size are cohesionless rubble-piles (see Fig. 4). Also Britt et al. [18], from an analysis of known asteroid densities, defines two categories of shattered objects among such rubble-pile asteroids. These authors distinguish the fractured or heavily shattered objects with porosity in the range 10-25% from the loosely consolidated rubble-piles with porosities in the range 30-80%. Last, a classification that considers the relative tensile strength together with the porosity has been proposed [19]. Since such a rubble-pile asteroid should be a cohesionless and gravity-dominated body, one can expect it to have a figure of equilibrium. Isaac Newton, back in 1687 in his *Principia*, derived the flattening ($\epsilon \equiv 1 - \frac{\text{polar radius}}{\text{equatorial radius}}$) of the Earth by considering it as a fluid of constant density and equalizing the weight of the water as due to gravitational and centrifugal acceleration in two radial canals, one directed toward the pole and another directed toward the equator. He could so explain from his theory of gravitation that the equator is not more submerged by the oceans than the poles. The flattening derived by Newton⁶ $\epsilon = 1/230$ is that of the rotating Earth considered as an incompressible fluid, yet different from the modern value 1/298.3. Following the work of Newton other mathematicians studied the figures of equilibrium of a rotating mass, the reader is referred to one of

⁶Ch. Huygens in 1690 had a very similar approach but fundamentally different in that he did not believe in Newton’s gravitation law and his calculation yield $\epsilon = 1/578$, i.e. as if all the mass of the Earth were concentrated in its center.

the most comprehensive work that was made by S. Chandrasekhar [20]. A few decades after the work of Newton, Maclaurin (1742) in England showed that ellipsoids of revolution are equilibrium figures of homogeneous mass of fluid in rotation, and A. Clairaut (1743) in France gave a relation between the density distribution inside the Earth and its flattening at its surface. This was followed one century later by the result of Jacobi (1834), who showed that there also exist a class of tri-axial ellipsoids for the figure of hydrostatic equilibrium. We will see three approaches to this problem considering different rheology: incompressible fluid, compressible material in the linear-elastic regime, and plastic-elastic material before yield.

4.1 Hydrostatic Equilibrium

For an incompressible fluid at rest to be in equilibrium the pressure and centrifugal force must balance the gravity. The equation of hydrostatic equilibrium states that, for each element of volume, the external body force and the boundary surface force must balance; so that considering the force per unit mass $\mathbf{f}$ and the pressure p we have $\rho \mathbf{f} = \nabla \mathbf{p}$. When the external force represents a scalar potential field $\mathbf{f} = \text{grad} \mathbf{U}$ and the density is constant one finds $\rho \mathbf{U} - \mathbf{p} = \mathbf{0}$, hence under gravity only $p = \frac{2}{3} \pi G \rho^2 (R_0^2 - r^2)$. Similarly, stating that each volume element of the fluid is at rest also results in an isotropic stress tensor $\sigma_{ij} = -p \delta_{ij}$ where $p = -\sigma_{ii}/3$ is the hydrostatic or mechanic pressure: i.e. all normal (and compressive) stresses are equal and the shearing stresses are zero.

Introducing uniform rotation along one axis, the force still is obtained from a potential, so that one sees from a $\nabla \mathbf{p} = \rho \nabla \mathbf{U}$ that the external surface must also be an equipotential surface, with equal density and pressure. This is a necessary condition for figures of hydrostatic equilibrium; to be sufficient the total energy (gravitational, kinetic, tidal, ...) has to be minimized. For instance, considering a flattened sphere of eccentricity e – together with fact that total mass M and angular momentum J are conserved – one finds the figure of equilibrium by minimizing the energy $E = W + T$ (sum of gravitational and kinetic energy) over the two free parameters that are e.g. the density ρ and the ellipticity e : $\partial E / \partial \rho = 0$ and $\partial E / \partial e = 0$. By introducing the (diagonal) potential-energy tensor:

$$\mathcal{U}_{ii} = \int_V \rho x_j \frac{\partial U}{\partial x_i} dx$$

and the inertia tensor $I_{ij} = \int_V \rho x_i x_j dx$, Chandrasekhar derives the virial relation in tensor form:

$$\mathcal{U}_{11} + \Omega^2 I_{11} = \mathcal{U}_{22} + \Omega^2 I_{22} = \mathcal{U}_{33}$$

For a homogeneous ellipsoid of semiaxis ($a_1 \geq a_2 \geq a_3$), spinning along its shortest axis, the gravitational potential and kinetic energies are respectively:

$$W = -\frac{3}{10} \frac{G M^2}{R^3} \sum_i A_i a_i^2 \quad \text{and} \quad T = \frac{5J^2}{2M(a_1^2 + a_2^2)}$$

where A_1, A_2 and A_3 are calculated in terms of Jacobi integrals involving the ellipsoid's ratio only (see e.g. [20]):

$$A_i = (a_2/a_1)(a_3/a_1) \int_0^\infty ((a_i/a_1)^2 + u)^{-1} \Delta^{-1} du; \quad i = [1..3]$$

$$\Delta = \sqrt{1 + u + ((a_2/a_1)^2 + u) + ((a_3/a_1)^2 + u)} \quad (6)$$

Note however that the adimensional A_i coefficients are defined differently to e.g. [21]; here we have $\sum_i A_i = 2$.

We only give a brief and general outline of the problem resolution and analysis of the configuration's dynamical and secular stability. The reader will find a very concise treatment in [20]. Here we are most interested in the figures the rotating fluid can take as a the result of the hydrostatic equilibrium hypothesis.

If there is no rotation, the equipotentials are spherical and so is the figure of equilibrium. A mass of homogeneous fluid with a relatively small ratio of rotational to gravitational potential energy $T/|W|$ will resemble a Maclaurin spheroid (Fig. 5), i.e. a axisymmetric spheroid with $a_1 = a_2 > a_3 > 0$. Then by putting the ellipticity:

$$e \equiv \sqrt{1 - (a_3/a_1)^2} \quad (7)$$

we have:

$$A_1 = A_2$$

$$A_1 = \frac{1}{e^3} \left(\sin^{-1} e - e \sqrt{1 - e^2} \right)$$

$$A_3 = \frac{1}{e^3} \left(\frac{e}{\sqrt{1 - e^2}} - \sin^{-1} e \right) \quad (8)$$

and the rotation frequency is given by:

$$\bar{\Omega}^2 \equiv \frac{\Omega^2}{\pi G \rho} = 2 \frac{\sqrt{1 - e^2}}{e^2} \left((3 - 2e^2) \frac{\sin^{-1} e}{e} - 3 \sqrt{1 - e^2} \right) \quad (9)$$

By increasing for instance the rotational energy, the fluid will evolve along this sequence through flatter configurations. At a sufficiently high $T/|W| = 0.1375$, and $\Omega^2/(\pi G \rho) = 0.374$, the equilibrium is secularly unstable, and one finds that the axisymmetric configuration is no longer the lowest energy state available. There is another sequence consisting of tri-axial ellipsoids ($a_1 > a_2 > a_3$) with some specific relation $\Phi(a_1/a_2, a_2/a_3) = 0$, the Jacobi sequence, along which a uniformly rotating incompressible fluid

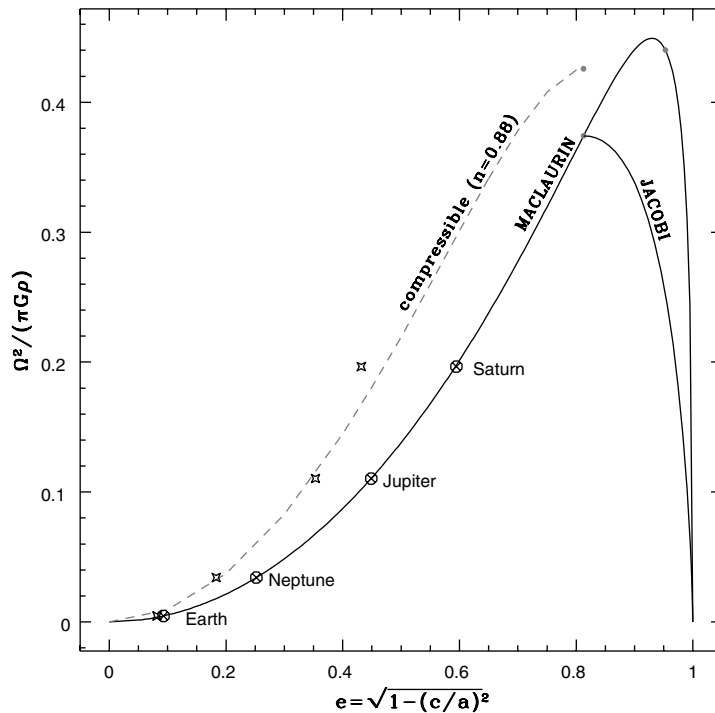


Fig. 5. Maclaurin and Jacobi sequences. Theoretical flattening of the homogeneous incompressible Earth, Jupiter, Saturn and Neptune (*circled crosses*). The observed flattening of the planets are lower (*crosses*). We also give the sequence of a compressible spheroid with a polytropic index of $n = 0.88$

will now evolve⁷. Another sequence at still larger $T/|W|$ that bifurcates from the Jacobi sequence at some instability is one yielding binary structures. The latter could explain the formation from a catastrophic collision of a binary asteroid system through rotational fission. We will focus on the result of Jacobi that is of particular interest here, since we know from asteroids lightcurves that their shapes are not ellipsoids of revolution but better approached by tri-axial ellipsoids. So, we can suggest that the shape of rubble-pile asteroids, as re-accumulation of a large number of aggregates with no internal cohesion, could mimic that of incompressible fluids in hydrostatic equilibrium and, depending on their angular momentum, could be Jacobi ellipsoid.

One possible application of such result is that knowing the shape and rotation period of an asteroid as derived from the lightcurve analysis, one can

⁷The case of viscous or non-uniformly rotating fluid, that can evolve along different ellipsoid sequences, is not discussed here.

determine – in the framework of an incompressible fluid – its bulk density⁸. This of course assumes that the asteroid shape is close to an ellipsoid and that its density is constant. We will give two illustrative examples. Let us first remind the relation on the shape of a Jacobi ellipsoid:

$$a_1^2 a_2^2 \frac{A_1 - A_2}{a_2^2 - a_1^2} = a_3^2 A_3$$

or equivalently

$$A_1 - (a_3/a_1)^2 A_3 = A_2 - (a_3/a_2)^2 A_3$$

Numerical results for Jacobi figures are given in 1 together with the corresponding rotational frequency sequence:

$$\bar{\Omega}^2 = \frac{\Omega^2}{\pi G \rho} = 2 (A_2 - (a_3/a_2)^2 A_3)$$

Table 1. Jacobi figures of hydrostatic equilibrium

a_2/a_1	a_3/a_2	$\bar{\Omega}$	a_2/a_1	a_3/a_2	$\bar{\Omega}$	a_2/a_1	a_3/a_2	$\bar{\Omega}$
1.00	0.583	0.3742	0.65	0.703	0.3475	0.30	0.865	0.2149
0.95	0.598	0.3738	0.60	0.723	0.3373	0.25	0.891	0.1813
0.90	0.613	0.3726	0.55	0.744	0.3248	0.20	0.918	0.1436
0.85	0.630	0.3703	0.50	0.767	0.3096	0.15	0.944	0.1027
0.80	0.647	0.3668	0.45	0.790	0.2913	0.10	0.968	0.0605
0.75	0.664	0.3620	0.40	0.814	0.2696	0.05	0.989	0.0219
0.70	0.683	0.3557	0.35	0.839	0.2443	0.00	$\rightarrow 1$	$\rightarrow 0$

The first example is 45 Eugenia, which body appears to have approximately the shape of a Jacobi ellipsoid ($a_1/a_2 \sim 1.35$; $a_2/a_3 \sim 1.5$). Knowing that its rotation period is $P = 5.7$ hr we find a bulk density of $\rho_b = 1.24$. Further, assuming that this S-type asteroid is constituted with material of density $\rho_g \simeq 2.3 - 3$, we find a macro-porosity of $\sim 45 - 60$ %. Such porosity also seems in good agreement with our rubble-pile hypothesis, and would roughly correspond to a random packing of aggregates. Moreover Eugenia is known to possess a satellite (would it be the outcome of a catastrophic collision, it would also enforce the rubble-pile hypothesis), that orbits its primary in ≈ 4.7 days, so that one can independently estimate the bulk density $\rho_b = M/V \simeq 1.2^{+0.6}_{-0.2}$ [22]. This is in good agreement with the value obtained from the hydrostatic fluid model. This suggests that Eugenia is a Jacobi ellipsoid, and its overall shape – on the macroscopic scale – follows equipotential

⁸Newton originally used this scheme in his *Principae* to derive the density of Jupiter

surface. Moreover Eugenia could be a homogeneous body with a constant density profile. The second example is 63 Ausonia which shape is well approached by an ellipsoid but neither a Maclaurin (oblate) spheroid nor a Jacobi ellipsoid. However, the shortest axis being not very well determined by present high resolution observations, let's assume that the shape of Ausonia is close to that of a Jacobi ellipsoid. Then the large flattening of this body ($a_1/a_2 \sim 2.2$) would provide a density of $\rho_b = 0.6$. Further, assuming this C-type asteroid is constituted of material with density $\rho_g \simeq 2 - 2.5$, we find a macro-porosity of $\sim 70 - 76\%$, which value seems rather unrealistic! On the other hand, the observed lightcurves amplitude of Ausonia could be obtained with a binary asteroid where each component, in hydrostatic equilibrium [23], would have a somewhat higher density [24]. We shall however see in Sect. 5.3 that Ausonia actually is a single body with a shape close to a prolate spheroid. Last, if one plots the observed asteroids shapes against the Maclaurin and Jacobi shapes (see Fig. 6), one clearly sees that asteroids shapes generally depart from equipotential surfaces.

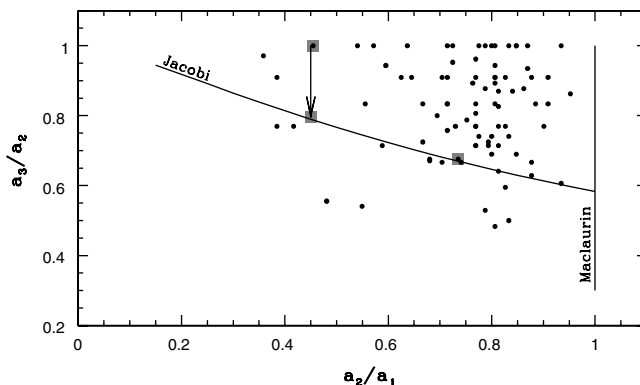


Fig. 6. Observed asteroids axis ratio against Maclaurin spheroids and Jacobi ellipsoids. Are overplotted the data for Eugenia on the right (*filled squares*), and on the left the data for the prolate spheroid Ausonia as well as the hypothetical Jacobi shape Ausonia

If Eugenia can be considered as a Jacobi ellipsoid, it seems nevertheless that, as seen before, asteroids in general and rubble piles in particular do not follow such figures of equilibrium for inviscid and incompressible fluids. Interesting, instead of the Jacobi ellipsoid shapes, the prolate spheroid ($a_1 > a_2 = a_3$) appear to be a more common shape among asteroids. The previous analysis can be completed, in the case of stars as well as in the case of solar system bodies, by considering – inviscid – compressible fluids. In this case the density and pressure are no more constant through the body, and are related by some general law $f(\rho, p, T) = 0$. For instance Laplace considered the relation $dp/d\rho = h\rho$, and Roche $dp/d\rho = h\rho + h'\rho^2$, h and h' being

constants. In a similar way one can also consider bodies of incompressible but non homogeneous fluids. In addition to the numerical experiments simulating the behavior of compressible stars, let us mention the analytical work achieved in [25] for describing the ellipsoidal figures of equilibrium in the compressible case. The authors have considered polytropes of index n , $p = K\rho^\Gamma$ with $\Gamma = 1 + 1/n$. Briefly, such fluid compressibility will change the shape of the sequence as can be seen in Fig. 5, but not the overall shape of the figure of equilibrium. In other words, Maclaurin spheroid and Jacobi ellipsoids are still figures of equilibrium but the rotational frequency for a given shape is different to that of the incompressible case. Moreover, as shown in Fig. 5, for a given shape the rotational frequency is larger when considering this particular density distribution; thus the bulk density for a given shape is smaller than it is in the incompressible case. From that it also appears that the knowledge of the rotation period and geometric flattening (or shape) is not sufficient to obtain information on the density distribution inside the body. Also such compressibility would not provide a more realistic bulk density for our Ausonia example.

4.2 Elastostatic Equilibrium and Elastic-Plastic Theories

Another approach of interest for solid bodies and that extends the approach of fluids is to consider elastic bodies⁹. We now consider a rheology where there is a linear relation between stress σ_{ij} and strain ϵ_{ij} , and the law of constraint is given by Hooke's law:

$$\sigma_{ij} = \lambda \epsilon_{kk} \delta_{ij} + 2\mu \epsilon_{ij}$$

where (λ, μ) are the two Lamé coefficients of elasticity. The strain is derived from the deformation field $u_i(x_i)$ in the frame $(\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$ by:

$$\epsilon_{ij} = \frac{1}{2} \left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i} \right)$$

and the strain-stress relation can also be written as (see e.g. [26]):

$$E \epsilon_{ij} = (1 + \nu) \sigma_{ij} - \nu \sum_k \sigma_{kk} \delta_{ij} \quad (10)$$

where $E > 0$ is Young's modulus, and $-1 < \nu < 1/2$ the Poisson¹⁰ ratio, coefficients that only depend on the considered material. For instance, in the case of a simple compression $\sigma_{11} < 0, \sigma_{22} = \sigma_{33} = 0$ along axis $\mathbf{x}_1$, one has $\epsilon_{11} = 1/E \sigma_{11}$, and the corresponding deformations rate along the perpendicular directions are $\epsilon_{22} = \epsilon_{33} = -\nu/E \sigma_{11} = -\nu \epsilon_{11}$. The couple of parameters (E, ν) is uniquely connected to the two Lamé coefficients. These coefficients

⁹And, say, thermoreologically simple bodies

¹⁰In practice we have $\nu > 0$ although negative Poisson's ratio have been witnessed in some foam [27], see URL <http://silver.neep.wisc.edu/~lakes/Poisson.html>

also express the mechanic dissipation inside the material. Considering by continuity the limiting case where $\nu \rightarrow 1/2$, the shearing-stiffness modulus is $\mu = 0$ and one should find the results for the incompressible and inviscid fluids case. The static equilibrium is obtained when the sum of internal and external body forces vanishes $\mathbf{div} \Sigma + \mathbf{f} = \mathbf{0}$, where $\Sigma = [\sigma_{ij}]$ is the stress tensor. Considering the gravitational and centrifugal potentials from which the forces are derived ($\mathbf{f} = \rho \mathbf{grad}(U)$), one obtains:

$$\sigma_{ij,j} = \frac{\partial \sigma_{ij}}{\partial x_j} = -f_i = \rho \frac{\partial (U_G + U_C)}{\partial x_i} \quad (11)$$

As seen above in Sect. 3 we can assume that the object is spinning along its shortest axis with constant angular rate Ω and write the centrifugal potential

$$U_C = \Omega^2 (x_1^2 + x_2^2)/2$$

The gravitational potential is given by:

$$U_G = \pi G \rho a_1^3 \left(U_0 - \sum_i A_i x_i^2 \right)$$

where the coefficients have been defined in (6), and $U_0 = \int_0^\infty \Delta^{-1} du$. Solution of equilibrium can next be obtained by minimization of the elastic energy. The problem is analytically tractable because the loads are linear in the spatial coordinates and because of the symmetries in the considered figure. For instance this approach has been applied to the non-rigid spheroidal Earth [28]. The analytical treatment in the case of tri-axial ellipsoids is more cumbersome and has been treated in e.g. [29–31]. Considering an homogeneous, isotropic, linear-elastic, and slightly compressible material with $\nu = 0.45; 0.499$, Washabaug & Scheeres [31] have shown that, at low angular momentum, ellipsoidal figures with compressive stresses at the surface exist (i.e. the presence of tensile strength is not needed) but they generally lie only in the vicinity the minimal energy state. They also showed that the elastic energy minima are weaker for compressible material, but nevertheless they occur at high angular momenta and for very elongated shapes which shapes are not observed among asteroids.

Eventually, we will discuss a more general approach that does not depend on the actual stress-strain behavior or possible residual stresses, but we will consider the *limit* stresses of an elastic-plastic body. Starting from (11) – and considering only ellipsoids – the equilibrium equations and boundary conditions leaves three degree of freedom in the general solution. Thus one can consider that the material is fluid (imposing that the shear stresses are vanishing and all normal stresses are equal), or one can consider linear-elastic isotropic material by introducing an additional relation between strain and stresses from Hooke’s law. These are the two options we have considered so far. Considering fluids or elastic deformations to model asteroids would not allow for large fails, boulders or craters on the surface of these small rocky

bodies as can be witnessed for instance on the surface of the planetary satellite Phobos or asteroid (253) Mathilde. It has been suggested that cohesionless bodies could maintain shapes significantly different from figures of hydrostatic equilibrium, so long as i) one assumes that the rotating mass is not a fluid but behaves as a granular soil with non negligible solid-to-solid friction, and ii) internal stresses are not high enough to crush individual particles [32, 33]. In this case the material can sustain non negligible tangential force before plastic flow. The Mohr-Coulomb criteria is generally used for deriving the maximal stress strain before yield of a given soil. The stress tensor $\Sigma = [\sigma_{ij}]$ being symmetric, it is diagonalizable; given the principal stresses $\sigma_1, \sigma_2, \sigma_3$ and neglecting the cohesive strength, the Mohr-Coulomb criteria states:

$$\tan \phi \geq \frac{\sigma_1/\sigma_3 - 1}{2\sqrt{\sigma_1/\sigma_3}}$$

that is, the maximal tangential stress is limited by only the most and least (compressive) normal stresses σ_1 and σ_3 . In such case we no more have an equality that provides the unique figure of equilibrium, but instead a range of possible configurations. Considering again tri-axial ellipsoids of constant density, it appears that the ratio σ_1/σ_3 does not depend on the spatial coordinates, and one can plot the rotational frequency versus the axis ratio for a given friction angle [33]. We see in Fig. 7 that the hydrostatic equilibrium is obtained for negligible friction, and for a given friction angle (e.g. $\Phi = 15^\circ$) there is a zone of possible ellipsoidal figures. That is, for those tri-axial ellipsoids the gravitational and centrifugal load are small enough to be sustained by the friction and hence avoid failure. We also see that the rotation frequency is limited at the largest possible friction $\Phi = 90^\circ$, in particular for a sphere one has [4]:

$$\sqrt{\Omega_{\max}} = 4\pi/3$$

but this does depend linearly on the axis ratio. Last if one plots the data from known asteroid, assuming typical densities in the range 1–3, it eventually appears that the great majority of observed shapes is consistent with such a cohesionless Mohr-Coulomb model and a friction angle of $\approx 25^\circ$, which value seems realistic and typical of dry terrestrial soils (it is $\approx 30^\circ$ for sand).

4.3 Binary Systems and the Density Profile

Asteroids do have satellites too! Since the first one that was discovered by the Galileo probe during its fly-by with Ida in 1993, about a dozen additional systems have been identified by ground-based observations (radar, adaptive optics, photometry) in the NEO or KBO population as well as in the main belt. Note that previous attempts from the Hubble Space Telescope imaging instrument were unsuccessful [34]. The presence of a satellite (generally 10-20 times smaller than the primary) is of high value to determine the mass of the primary from Kepler's third law and eventually its bulk density. Moreover,

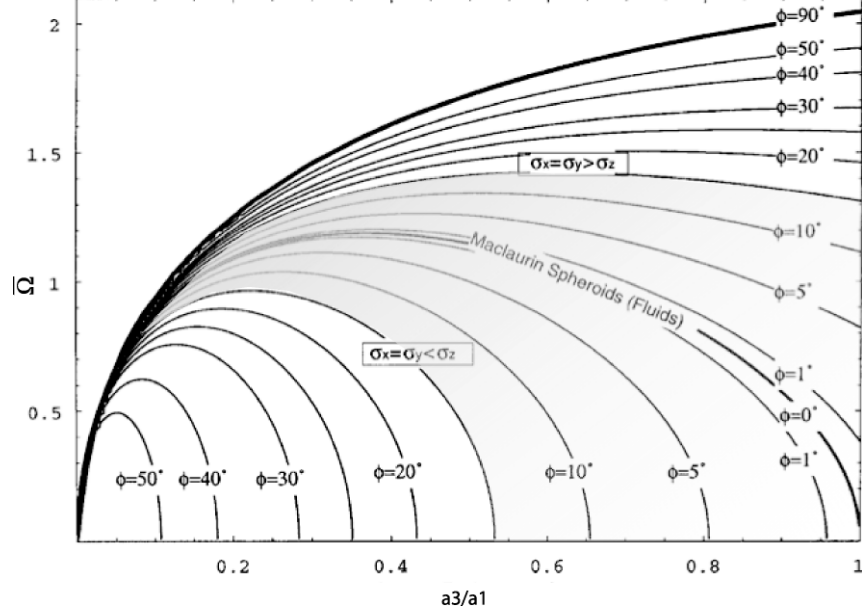


Fig. 7. Limit rotational frequency figures for oblate spheroids (from [33])

since these objects are rather flattened or elongated, the dynamical perturbations due to the non-spherical gravitational potential are expected to be relatively large. As seen previously the (secular) perturbations depends on the potential, or mass distribution [35]. If the spin rate of the primary is much larger than the orbital rate of the satellite ($\Omega_{\text{rot}} \gg n$), the secular effect of the C_{22} is negligible, and the major perturbation arises from the J_2 coefficient:

$$J_2 = -C_{20} = -\frac{I_1 + I_2 - 2I_3}{2Ma_e^2}$$

Further, the secular terms are obtained from Lagrange equations:

$$\begin{aligned} \dot{a} &= \dot{e} = \dot{i} = 0 \\ \dot{\omega} &= \frac{3}{4} \frac{n a_e^2}{a^2 (1 - e^2)^2} (1 - 5 \cos^2 i) C_{20} + \dots \\ \dot{\Omega} &= \frac{3}{2} \frac{n a_e^2}{a^2 (1 - e^2)^2} \cos i C_{20} + \dots \\ \dot{M} &= \frac{3}{4} \frac{n a_e^2}{a^2 (1 - e^2)^{3/2}} (1 - 3 \cos^2 i) C_{20} + \dots \end{aligned} \quad (12)$$

We know since Clairaut and Radau that the knowledge of the dynamical and geometric flattening brings insight on the density distribution inside the body. Thus observing the secular perturbation on the satellite's orbit over

successive months provides the dynamical flattening while observing photometric variation over successive apparitions provides the geometric flattening. By comparing both one can at least test the hypothesis of constant bulk density inside the (primary) asteroid. This has been applied to the orbit of Kalliope's companion [36]. Assuming a homogeneous primary with constant density distribution the observed geometrical flattening provides a dynamical $J_2 = \frac{1}{10} \frac{a_1^2 + a_2^2 - 2a_3^2}{a_e^2}$ and in turn a precession rate of $\dot{\omega} \sim 0.3 \text{ deg/day}$. This value is in severe conflict with the observed one $\approx 0.7 \text{ deg/day}$. As shown in Fig. 8 there are three possible explanations that maybe all concur together to this discrepancy: the size of the body, the geometrical flattening, and the density distribution. Asteroids diameters are essentially given by an indirect method from observations with the IRAS satellite, and can be in error by 10% or more [37]. The shape being derived from disk-integrated photometric data and not high resolution imaging is not known with the best accuracy neither. Last, including an empirical but simple density distribution of the form $\rho(r, \theta, \phi) = \rho_0 r^\alpha$; $\alpha \in \mathbb{R}$, i.e. which *increases* as we progress toward the surface ($r = 1$), one can write the zonal harmonic as a function of the one for $\alpha = 0$, i.e. at constant density J_2^0 :

$$J_2 = \frac{5\alpha + 3}{3\alpha + 5} J_2^0$$

Hence for a given shape, the pericenter precession is increased by concentrating the mass at the outer surface of the body. Assuming that the large precession is due to the non homogeneous mass distribution alone, one finds a density at Kalliope's surface of ≈ 7 , which is marginally acceptable. Let's now consider that Kalliope is a size-sorted rubble-pile, where the larger and more irregular rocks are in the central part and the smaller material is kept by friction at the outer layers [38]. Such mass distribution could correspond to a body of homogeneous but size-sorted material with larger relative voids (or porosity) in the central part, and more densely packed material toward the surface.

5 The Determination of Shape and Spin Parameters by Hubble Space Telescope

5.1 The FGS Interferometer

The Hubble Space Telescope (HST) is a complex of instruments, built around a 2.4-meters telescope, orbiting the Earth. For several reasons (pressure on observing time requests, time constraints due to the orbit, complexity of the instruments, etc.) it is sometimes very difficult to apply for successfully and to use.

However, its great advantage – being outside the Earth atmosphere – allows to obtain otherwise difficult measurements and observations. While its

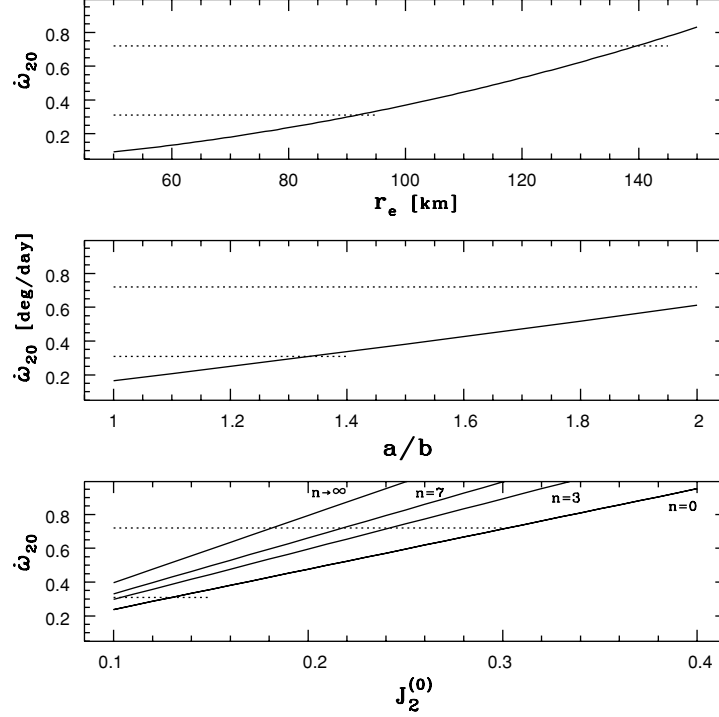


Fig. 8. Secular periastron advance $\dot{\omega}_{20}$ for the orbit of 22 Kalliope's companion

imaging capabilities of deep-sky objects and planetary surfaces/atmospheres are well known, they are not sufficient to perform accurate measurements of asteroid shapes and sizes. In fact, the highest resolution is reached by the Planetary Camera, having a plate scale of 46 mas/pixel¹¹. This value is of the same order than the apparent size of several, interesting main belt objects at opposition, and allows some resolution to be achieved on a very restricted set of the largest bodies only.

A much more sensitive instrument is the Fine Guidance Sensor (FGS), an interferometer normally used to allow careful pointing and guiding of the HST while imaging is performed by the main CCD cameras. Three FGS instruments are mounted close to the focal plane, and each of them works by producing interference between the two beams coming from the defocalised semi-pupils of the telescope. Inside the FGS the beam is divided into two parts, associated to two perpendicular axis (in the following we refer to them as the FGS-X and FGS-Y axis). Each beam enters a Koester prism. Inside this device the self-interference occurs. The resulting two beams exiting a single Koester are then collected by two photometers, measuring their flux. The important feature

¹¹In the following we express all apparent sizes in milli-arcseconds (mas) = 10^{-3} seconds of arc

to note here, is that the difference between the fluxes is a function of the inclination of the incoming wavefront¹².

The FGS is able to ‘scan’ the focal plane in straight line, along the FGS-X and FGS-Y axis, in steps of 1–2 mas, and over a few arcseconds. Each step corresponds to a different inclination of the wavefront and thus to a different response of the photometers. A response function, called “S-curve”, is reconstructed from the flux difference normalized to the total flux.

It is clear, then, that the response for an extended source depends upon the light distribution on the focal plane. The FGS sensitivity being optimal for spatial frequencies ≤ 200 mas, it is normally employed to measure the diameter of large star disks or the separation of close doubles. An example of the simulated response curve for an extended, uniform disk, or for a double disk, is given in fig. (Fig. 9). As it can be seen, the largest is the source, the smallest is the amplitude of the S-curve. A double object has the tendency to double the curve peaks. In order to better understand the results presented in the following, it must be noted that each scan by the FGS produces two S-curves, one for each axis. As a consequence, the extracted informations concern the size of the studied object as projected along the two FGS axis.

In recent years, the FGS has been used for the first time to measure rapidly rotating and moving objects: the asteroids. Six objects having interesting spin and lightcurve properties have been chosen, each observed for an HST “orbit”, i.e. for a total duration of about 40 minutes (corresponding to 30-40 FGS scans). During this time lapse two movements affect the observation: the spin of the object and its proper motion. The first one is highly desirable, since it can help to constrain the asteroid shape by studying the variations of the FGS-X and FGS-Y projections. The second one, on the other hand, has been minimized by choosing epochs close to the stationary points in the asteroids apparent motion, and by applying a correction (a posteriori) to compensate for the parallax due to HST displacement.

In summary, each set of scans for each asteroid contains informations on its shape and size, as projected over FGS-X and FGS-Y axis, over a limited fraction of the rotation period – roughly corresponding, for the given set of objects, to about the 10% (about 30-40 degrees of rotation). It can thus be supposed that a fitting of a model to the set of S-curves can lead to the reconstruction of sizes and shapes. The degree of accuracy (i.e. the number of parameters to be determined) will depend upon the quantity of observations and their signal/noise ratio.

5.2 From Data to Modeling

In general, given a set of FGS-X and FGS-Y S-curve couples at each epoch, the fitting procedure begins by determining shape, orientation and size of the on-sky projected ellipse that separately accommodates best to each curve.

¹²More details can be learned by browsing the HST/FGS on-line handbook at URL: <http://www.stsci.edu/instruments/fgs/>.

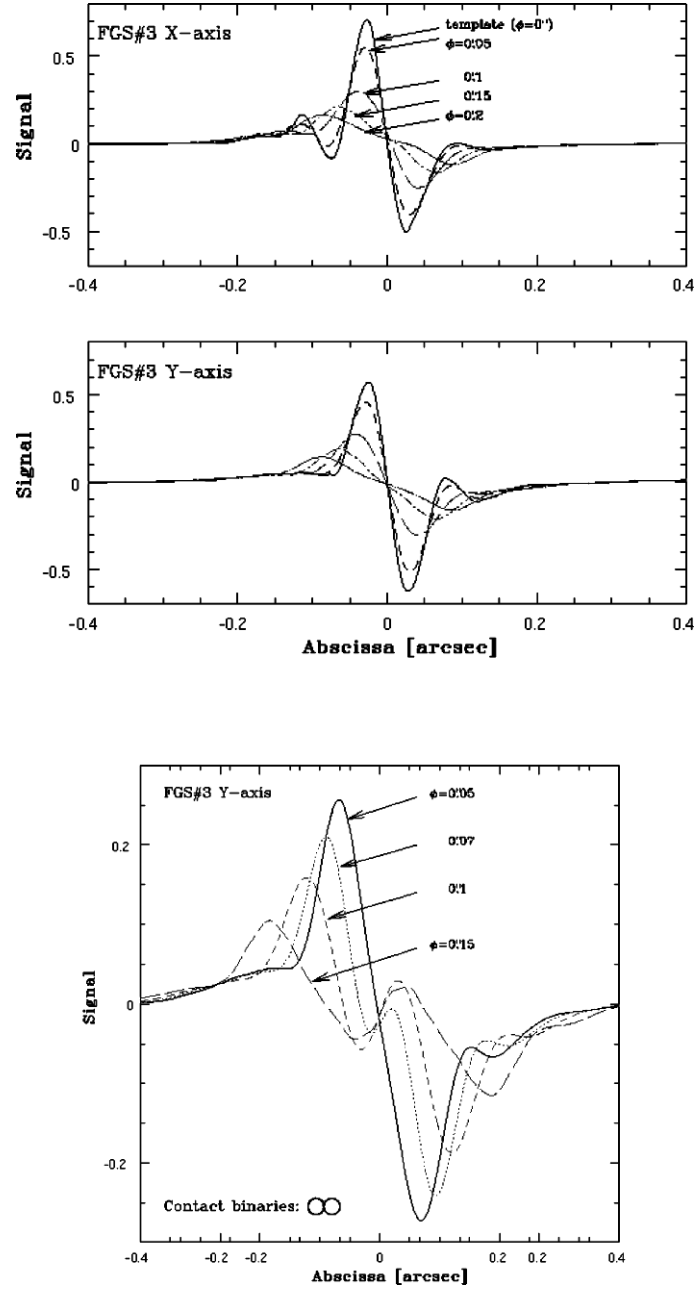


Fig. 9. *Upper panels:* the simulated response curve, along the FGS-X and FGS-Y axis, for a uniform, circular source of different sizes (indicated in arcsec). The “template” curve correspond to the observation of a point-like star. *Lower panel:* the simulated response curve for a double source composed by two equal, tangent disks of different diameters

The problem, then, is to identify which is the solid, three-dimensional body that is capable of producing the observed, projected ellipses. If a traditional equilibrium figure is searched for, an ellipsoid with a given orientation of the spin axis will be the figure of choice.

It must be noted that traditional photometry from Earth-based telescopes normally records brightness variations associated to the object shape, directly yielding its rotation period. From lightcurves taken at different epochs, a first estimation of the spin axis direction can be obtained. In general, however, some symmetries in the problem does not allow to discriminate between degenerate solutions. The result of photometric pole determination is thus expressed by two possible spin axis directions, each yielding, for a given epoch, the same object area projected on the sky, i.e. the same brightness. However, an instrument capable of directly detecting the orientation of the shape and its variation in time, can immediately eliminate the ambiguity and help to discard one of the two solutions. This is the first result that have been derived by HST observations.

Thus, having selected the good pole solution between the pair available, the model is fitted to the S-curve is that of an ellipsoid of uniform brightness. The shape is described by the three ellipsoid semi-axis a , b and c . Since the uncertainty in pole coordinates can reach several degrees, a trial and error adjustment, reducing the O-C, is performed. In fact, we realized that additionally solving the fit for pole coordinates seems not to add significant improvements given the set of data currently available. Thus, pole coordinates do not enter the core of the fit process.

After having determined the best-fitting ellipsoid, we will have a set of values corresponding to semi-axis sizes, spin pole coordinates and rotational phase. In turn, this allows to re-compute for each epoch the ellipse projected on the sky. Another iteration can thus be performed repeating the whole process from the beginning. If an ellipsoidal shape solution exists, the parameters rapidly converges and the residuals collapse. Significant residuals are, on the contrary, the indication of some detectable departure from the simple assumption on the shape, and some more complex models deserve to be taken into account.

5.3 Some Significant Examples

In the following we illustrate some results that we think to be particularly significant. The accuracy of the results, their limitations and the practical difficulties should be apparent. A complete review of results is published in [39–41], while the following table reviews the main parameters derived.

63 Ausonia

63 Ausonia was one of the brightest objects in the set, and the first to be observed. The signal-to-noise ratio is rather good, so the fitting process operates in ideal conditions. Figure 10 shows the S-curves of (63) Ausonia for a

Table 2. A summary of FGS size measurements, obtained by considering ellipsoidal models. (216) Kleopatra, having a complex shape, is not shown here. The last column gives the ratio of the axis; the parentheses indicate that either b or c are not well constrained.

Name	a, b, c [km]	a/b	a/c
(15) Eunomia	181, 103, 102	(1.76)	1.78
(43) Ariadne	45, 26, 26	1.71	(1.71)
(44) Nysa	59, 35, 35	(1.72)	1.72
(63) Ausonia	75, 33, 33	2.28	(2.28)
(624) Hektor	62, 28, 28	2.21	2.21

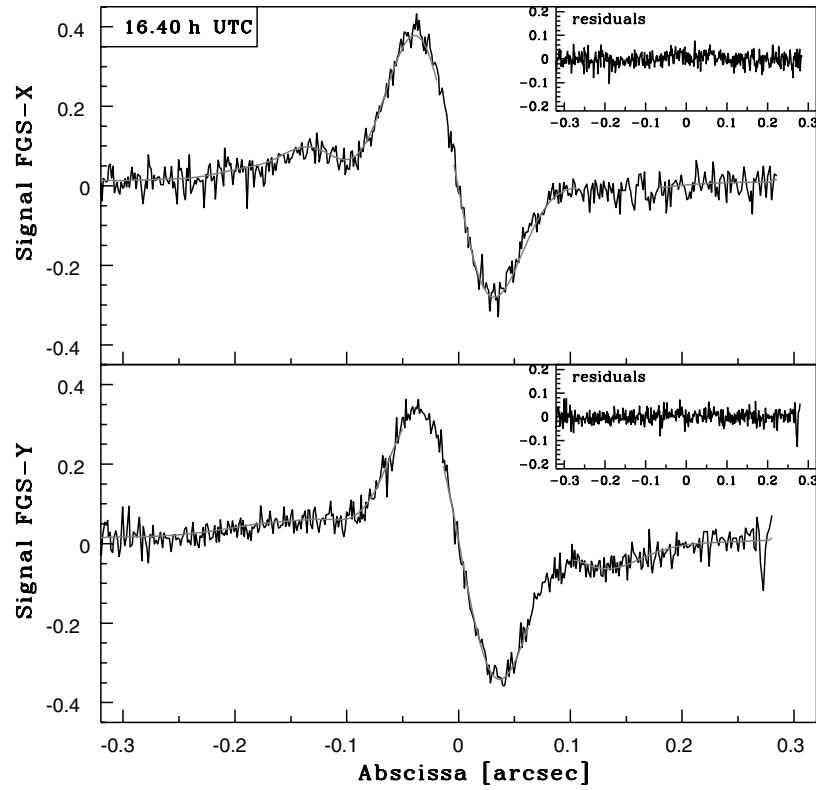


Fig. 10. The Ausonia S-curves at the beginning of the observation, fitted by using the tri-axial ellipsoid given in Table 2. The residuals of the fit given in the inset are very small.

selected epoch. The residuals of the fit with a three axis ellipsoid are shown in the inset and are very small in this case, showing that the identified 3D ellipsoid is completely consistent with the available FGS data, as discussed above.

Due to its orientation on the sky, the asteroid, while rotating, exhibits some interesting variations in the length of the projected a and b axis, while c (coinciding with the rotation axis) does not move and its influence on the FGS measurements is minor. For this reason, the value of the c axis is affected by a high uncertainty (around 5 mas or more) while a and b are constrained to a 1 mas level. These figures can be considered to be typical for this kind of model-dependent fit.

216 Kleopatra

This was most irregular body observed. A little time before the HST observations, it was observed by radar [42]. The reconstructed shape seemed to hint to a bi-lobated object, very elongated and irregular.

The HST/FGS signal suggests an elongated shape, well approximated by two – not detached – ellipsoids, whose signature is well visible in the S-curve. The overall shape appears to be more elongated and flattened in comparison to radar data. Details are given in [40] and [43].

624 Hektor

This Trojan asteroids is the faintest asteroid observed by HST/FGS during the program ($V=15.0$ at the epoch of the observation), being at an average distance corresponding to the semi-major axis of the orbit of Jupiter. Due to its faintness, the S/N ratio is very small and it probably represents the lowest possible for this kind of studies by the FGS. The response functions is best fitted by a very elongated shape, but due to the low S/N ratio it is not possible to clarify if (624) Hektor can really be considered to be a contact binary as supposed in the past by Weidenschilling [44].

Figure 12 shows the best-fit shape obtained by HST observations.

15 Eunomia

Together with the previous object, (15) Eunomia is another difficult target. (15) Eunomia is the object having the largest apparent size, as shown by the small amplitude of the S-curve. The shape of the signal is consistent with a shape more complex than a simple ellipsoid, as shown in Fig. 13. Unfortunately, in order to better constrain its shape, more observations and a widest coverage of its rotation would be necessary.

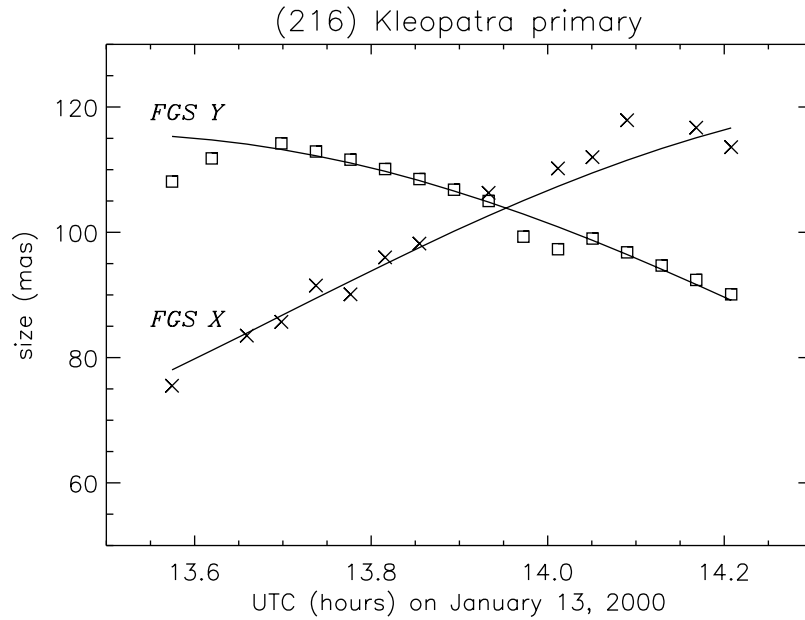


Fig. 11. The variation of the sizes projected on FGS-X and FGS-Y for the main component of the double-lobed asteroid (216) Kleopatra, during the observation. The shape of the object and its orientation relatively to the FGS axis are represented by the two ellipsoids below the curves

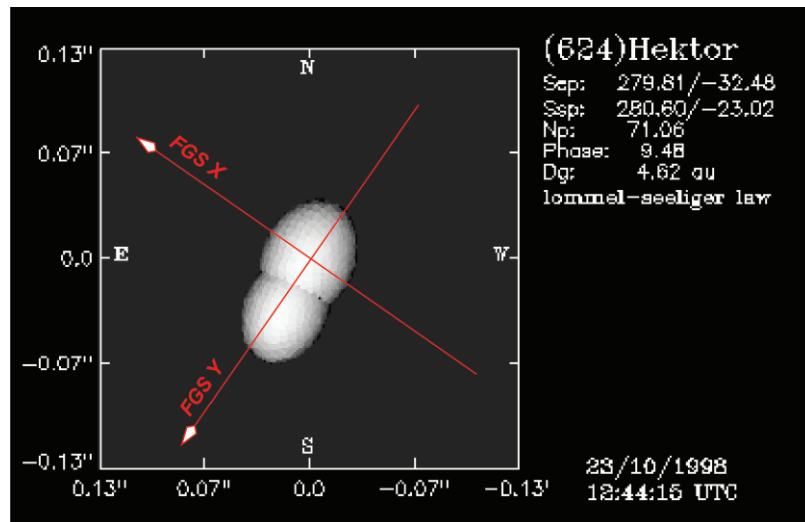


Fig. 12. The indicative shape of Hektor obtained by a best-fit of the FGS observations. The noisy data set does not allow to draw definitive conclusions.

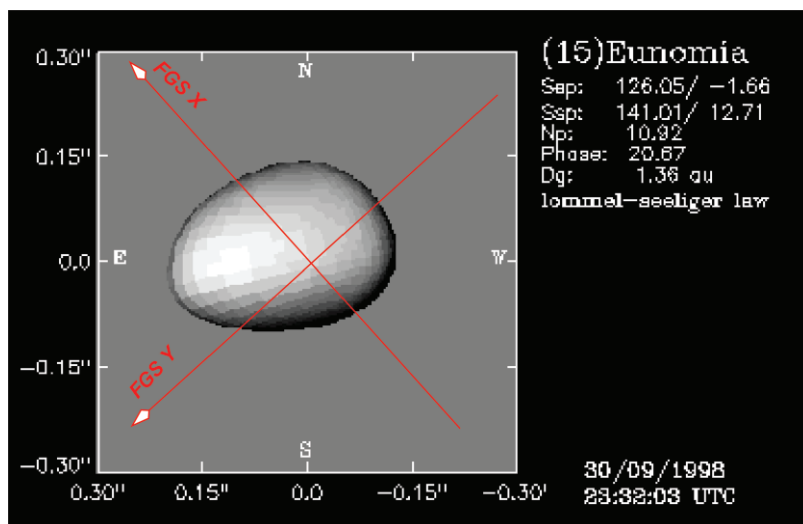


Fig. 13. *Lower panel:* the suggested shape of (15) Eunomia is consistent with the strongly asymmetric S-curve obtained for this object. Other possible solutions require the use of other constraints, such as those coming from photometry.

6 Conclusions

The observations by HST/FGS are a sensitive and powerful method to determine size and shape of asteroids. However, some limitations should be kept in mind.

First of all, the observing time is very limited and hard to obtain. For this reason, it is very difficult to sample the whole rotation curve of an object. Unfortunately, this is the only way to constrain complex shapes.

Furthermore, the shape solution that is found depends upon some *a priori* choices, such as the basic shape model, its scattering properties, the absence of albedo markings, etc. This details, while not changing the qualitative interpretation of the data, are probably important in order to define the ultimate precision of the observations.

Finally, the FGS can work on a limited sample of extended and bright asteroids. However, the same basic approach, coupled with more sophisticated data inversion techniques, can be applied to other optical interferometers, based on the Earth surface. We can thus hope that, in the future, the sample of asteroids for which size, shape and orientations are known will give us a more complete view of their physical properties. Last, combining photometric data to high resolution data and astrometric positions of asteroid satellite should bring insight of the primary asteroid interior and possible collisional history.

References

1. H. N. Russel. On the light-variations of asteroids and satellites. *Astrophysical Journal*, 24:1–18, 1906.
2. P. Magnusson, M. A. Barucci, J. D. Drummond, K. Lumme, and S. J. Ostro. Determination of pole orientations and shapes of asteroids. In *Asteroids II*, pp. 67–97, 1989.
3. M. Kaasalainen, J. Torppa, and J. Piironen. Models of Twenty Asteroids from Photometric Data. *Icarus*, 159:369–395, October 2002.
4. A. W. Harris. The Rotation Rates of Very Small Asteroids: Evidence for ‘Rubble Pile’ Structure. In *Lunar and Planetary Science Conference*, volume 27, pp. 493–494, March 1996.
5. M. Kaasalainen, S. Mottola, and M. Fulchignoni. Asteroid models from disk-integrated data. In *Asteroids III*, W. F. Bottke Jr., A. Cellino, P. Paolicchi, and R. P. Binzel (eds), University of Arizona Press, Tucson, pp. 139–150, 2002.
6. T. Tokieda. Spinning Bodies: A Tutorial. In: Lect. Notes Phys, **682** (2005), pp. 1–22.
7. L. Landau and E. Lifchitz. *Mécanique*. Editions en langues étrangères, Moscou, 1960.
8. P. Paolicchi, J. A. Burns, and S. J. Weidenschilling. Side Effects of Collisions: Spin Rate Changes, Tumbling Rotation States, and Binary Asteroids. In *Asteroids III*, W. F. Bottke Jr., A. Cellino, P. Paolicchi, and R. P. Binzel (eds), University of Arizona Press, Tucson, pp. 517–526, 2002.
9. N. H. Samarasinha and M. F. A’Hearn. Observational and dynamical constraints on the rotation of Comet P/Halley. *Icarus*, 93:194–225, October 1991.
10. S. J. Ostro, R. S. Hudson, K. D. Rosema, J. D. Giorgini, R. F. Jurgens, D. K. Yeomans, P. W. Chodas, R. Winkler, R. Rose, D. Choate, R. A. Cormier, D. Kelley, R. Littlefair, L. A. M. Benner, M. L. Thomas, and M. A. Slade. Asteroid 4179 Toutatis: 1996 Radar Observations. *Icarus*, 137:122–139, January 1999.
11. J. A. Burns and V. S. Safronov. Asteroid nutation angles. *M.N.R.A.S.*, 165:403–411, 1973.
12. A. W. Harris. Tumbling asteroids. *Icarus*, 107:209–211, January 1994.
13. M. Efroimsky. Euler, Jacobi, and missions to comets and asteroids. *Advances in Space Research*, 29:725–734, 2002.
14. P. Michel. Modelling Collisions Between Asteroids: From Laboratory Experiments to Numerical Simulations. In: Lect. Notes Phys. **682** (2005), pp. 127–153.
15. J. A. Burns. The angular momenta of solar system bodies – Implications for asteroid strengths. *Icarus*, 25:545–554, August 1975.
16. P. Farinella, P. Paolicchi, and V. Zappalà. The asteroids as outcomes of catastrophic collisions. *Icarus*, 52:409–433, December 1982.
17. P. Pravec and A. W. Harris. Fast and Slow Rotation of Asteroids. *Icarus*, 148:12–20, November 2000.
18. D. T. Britt, D. Yeomans, K. Housen, and G. Consolmagno. Asteroid Density, Porosity, and Structure. In *Asteroids III*, W. F. Bottke Jr., A. Cellino, P. Paolicchi, and R. P. Binzel (eds), University of Arizona Press, Tucson, pp. 485–500, 2002.
19. D.C. Richardson, Z.M. Leinhardt, H.J. Melosh, W.F. Bottke Jr., and E. Asphaug. Gravitational Aggregates: Evidence and Evolution. In *Asteroids III*, W. F. Bottke Jr., A. Cellino, P. Paolicchi, and R. P. Binzel (eds), University of Arizona Press, Tucson, pp. 501–515, 2002.

20. S. Chandrasekhar. *Ellipsoidal figures of equilibrium*. New York: Dover, 1987., 1987.
21. S. Chandrasekhar and N. R. Lebovitz. The Potentials and the Superpotentials of Homogeneous Ellipsoids. *Astrophysical Journal*, 136:1037–1047, November 1962.
22. W. J. Merline, L. M. Close, C. Dumas, C. R. Chapman, F. Roddier, F. Menard, D. C. Slater, G. Duvert, C. Shelton, and T. Morgan. Discovery of a moon orbiting the asteroid 45 eugenia. *Nature*, 401:565–467, October 1999.
23. G. Leone, P. Paolicchi, P. Farinella, and V. Zappalà. Equilibrium models of binary asteroids. *Astronomy & Astrophysics*, 140:265–272, November 1984.
24. A. Cellino, R. Pannunzio, V. Zappalà, P. Farinella, and P. Paolicchi. Do we observe light curves of binary asteroids? *Astronomy & Astrophysics*, 144:355–362, March 1985.
25. D. Lai, F. A. Rasio, and S. L. Shapiro. Ellipsoidal figures of equilibrium – compressible models. *Astrophysical Journal Supplement*, 88:205–252, September 1993.
26. P. Germain and P. Muller. *Introduction à la mécanique des milieux continus*. Masson, Paris, 1994.
27. R. S. Lakes. Foam structures with a negative Poisson’s ratio. *Science*, 235:1038–1040, 1987.
28. A. E. H. Love. The yielding of the Earth to disturbing forces. *M.N.R.A.S.*, 69:476–479, April 1909.
29. C. Chree. The equilibrium of an isotropic elastic solid ellipsoid. . . *Quart. J. Pure Appl. Math.*, 27:338–353, 1895.
30. A. R. Dobrovolskis. Internal stresses in Phobos and other triaxial bodies. *Icarus*, 52:136–148, October 1982.
31. P. D. Washabaugh and D. J. Scheeres. Energy and stress distributions in ellipsoids. *Icarus*, 159:314–321, October 2002.
32. M. Hammergren. Asteroid Shapes and the Angle of Repose. In *American Astronomical Society Meeting*, volume 194, page 1506, May 1999.
33. K. A. Holsapple. Equilibrium Configurations of Solid Cohesionless Bodies. *Icarus*, 154:432–448, December 2001.
34. A. Storrs, B. Weiss, B. Zellner, W. Burleson, R. Sichert, E. Wells, C. Kowal, and D. Tholen. Imaging observations of asteroids with hubble space telescope. *Icarus*, 137:260–268, February 1999.
35. L. Deleflie and Exertier P. Modelling and Characterizing the Earth’s Gravity Field: From Basic Principles to Current Purposes. In: *Lect. Notes Phys.* **682** (2005), pp. 76–97.
36. F. Marchis, P. Descamps, D. Hestroffer, J. Berthier, F. Vachier, A. Boccaletti, I. De Pater, and D. Gavel. A three-dimensional solution for the orbit of the asteroidal satellite of 22 Kalliope. *Icarus*, 165:112–120, 2003.
37. D. Hestroffer, P. Tanga, J. Berthier, A. Cellino, M. G. Lattanzi, M. di Martino, and V. Zappalà. HST/FGS interferometric observations of asteroids. *Mem. Soc. Astron. Italiana*, 73:674–678, 2002.
38. D. T. Britt and G. J. Consolmagno S. J. Modeling the Structure of High Porosity Asteroids. *Icarus*, 152:134–139, July 2001.
39. D. Hestroffer, P. Tanga, A. Cellino, F. Guglielmetti, M. G. Lattanzi, M. Di Martino, V. Zappalà, and J. Berthier. Asteroids observations with the Hubble Space Telescope. I. Observing strategy, and data analysis and modeling process. *Astronomy & Astrophysics*, 391:1123–1132, September 2002.

40. P. Tanga, D. Hestroffer, J. Berthier, A. Cellino, M. G. Lattanzi, M. di Martino, and V. Zappalà. NOTE: HST/FGS Observations of the Asteroid (216) Kleopatra. *Icarus*, 153:451–454, October 2001.
41. P. Tanga, D. Hestroffer, A. Cellino, M. Lattanzi, M. Di Martino, and V. Zappalà. Asteroid observations with the Hubble Space Telescope. II. Duplicity search and size measurements for 6 asteroids. *Astronomy & Astrophysics*, 401:733–741, April 2003.
42. S. J. Ostro, R. S. Hudson, M. C. Nolan, J.-L. Margot, D. J. Scheeres, D. B. Campbell, C. Magri, J. D. Giorgini, and D. K. Yeomans. Radar observations of asteroid 216 kleopatra. *Science*, 288:836–839, May 2000.
43. D. Hestroffer, J. Berthier, P. Descamps, P. Tanga, A. Cellino, M. Lattanzi, M. Di Martino, and V. Zappalà. Asteroid (216) Kleopatra. Tests of the radar-derived shape model. *Astronomy & Astrophysics*, 392:729–733, September 2002.
44. S. J. Weidenschilling. Hektor – nature and origin of a binary asteroid. *Icarus*, 44:807–809, December 1980.

Modelling Collisions Between Asteroids: From Laboratory Experiments to Numerical Simulations

Patrick Michel

Observatoire de la Côte d’Azur, UMR 6529 Cassini/CNRS, B.P. 4229, 06304 Nice
cedex 4, France

Patrick.michel@obs-azur.fr

Abstract. Thanks to the development of sophisticated numerical codes, a major breakthrough has been achieved in our understanding of the process involved in small body collisions. Such events play a fundamental role in all the stages of the formation and evolution of planetary systems, and more particularly of our Solar System. Laboratory experiments on centimeter-sized targets have been performed to improve our knowledge on this process, but their extrapolation to asteroid scales remains confronted to major difficulties. In this lecture, we present a brief review of our current understanding of the fragmentation process of solid bodies and its implementation in numerical codes aimed at simulating asteroid break-up events. The most recent results provided by numerical simulations are also presented. Although our current understanding is still based on several limitations and assumptions, the development of sophisticated numerical codes accounting for the fragmentation of an asteroid and for the gravitational interactions of the generated fragments have allowed to improve greatly our knowledge on the main mechanisms that are at the origin of some observed features in the asteroid belt. In particular, the simulations have demonstrated that, for bodies larger than several kilometers, the collisional process does not only involves the fragmentation of the asteroid but also the gravitational interactions between the fragments that are ejected. This latter mechanism can lead to the formation of large aggregates by gravitational reaccumulation of smaller fragments, allowing to explain the presence of large members within asteroid families. Numerical simulations of the complete process have thus been able to reproduce for the first time the main properties of asteroid families, each formed by the disruption of a large parent body, and also to derive some information on the possible internal structure of the parent bodies. A large amount of work remains however necessary to understand in deeper details the physical process as a function of material properties that are relevant to asteroids and to determine in a more quantitative way the outcome properties such as fragments’ shapes and rotational states.

1 Introduction

Collisional events are of great importance in the formation and evolution of planetary systems, including our own Solar System. In the first stages of planetary formation, low-velocity collisions between planetesimals drive planetary growth by collisional accretion. In the particular case of our Solar System, some energetic events also started to take place quite early, as indicated by the iron-rich composition of the planet Mercury which can be explained by the ejection of its original mantle due to a giant impact [5]. The Moon of our Earth is also understood as the product of ejected debris reaccumulation resulting from the impact of a planetesimal with our proto-Earth (e.g. [53]). In further stages, once the planets were formed, relative velocities between small bodies increased as a result of planetary perturbations. Consequently, our Solar System entered in a new regime of high impact energy, in which it continues to evolve currently. In this regime, collisions do not lead to accretion phenomena anymore but rather to disruptive events. In July 1994, observations of the impact of the comet Shoemaker Levy 9 on Jupiter have definitely proven that collisions continue to occur in our days. Furthermore, meteorites collected on Earth are another indication of the collisional activity as they are the remnants of collisions that take place in the asteroid Main Belt between the orbits of Mars and Jupiter. In this sense, collisions can be seen as representing an important threat against human efforts in space, which can even lead to the destruction of our biosphere. The collisional process is therefore not a second-order problem in the understanding of the past, present and future history of our Solar System; it is actually at the heart of its formation and evolution.

The scales of the phenomena that are involved in planetary and small body impacts are by far much larger than those reached in laboratory experiments. Extrapolations by 15 orders of magnitude in mass are necessary to achieve ranges that are relevant to asteroids and planetesimals. Theoretical models of catastrophic collisions try to fill this gap by establishing adimensional relationships between the projectile's size, the impact velocity, the target's strength, its density etc. that are supposed to be valid at all scales, and which are regrouped in the so-called *Scaling Laws* (see e.g. [22]). Nevertheless, such relationships are necessarily idealized, as they assume a uniformity of the process as well as a structural continuity. Consequently, they cannot predict with a high degree of reliability large-scale impact outcomes.

Numerical simulations remain the best approach to study the collisional process. It is now possible to simulate an impact with a certain degree of sophistication and reasonable accuracy thanks to dedicated numerical codes that have been developed recently, accompanied with the improvement of computer performances. Important problems can now be addressed concerning the physical nature of individual objects with a collisional history, the origin of asteroid families, the formation of planets through collisional accretion, etc. Impact experiments in laboratory remain useful to validate those numerical

models at small scales before they are applied to large-scale events. In the 1990s, an hydrodynamical code (hydrocode) based on the (*Smooth Particle Hydrodynamics*) method (SPH) has been developed by Benz and Asphaug ([6, 7]) who also included a model of brittle failure. It has successfully reproduced the results of centimeter-scale laboratory experiments. However, in the size range adapted to Solar System small bodies (>100 meters), its sole use is not sufficient as it is limited to the computation of the fragmentation of a solid body, whereas at those scales the role of gravity can strongly influence the collisional outcome. Indeed, it has been realized that in a collision involving large bodies, ejected fragments produced by the sole fragmentation process can actually interact gravitationally. Therefore, some reaccumulations can occur when relative velocities between fragments are below their mutual escape velocity and can eventually lead to a distribution of large aggregates. Accounting only for the fragmentation phase and neglecting the gravitational one would prevent the formation of such bodies and would then lead to different outcome properties that may not correspond to real ones.

To determine whether both our understanding of the collisional process and our methods to simulate it are correct, we have at our disposal a unique laboratory at the appropriate scales: asteroid families. More than 20 asteroid families have been identified in the asteroid belt, each corresponding to a group of small bodies sharing the same spectral and dynamical properties (see e.g., [24, 56]). This similarity of properties has suggested that all members within a group belonged to a larger asteroids, called the parent body, who has been catastrophically disrupted by an impact with a smaller projectile at high velocity. Therefore, each asteroid family constitutes the outcome properties of a collisional event at large scales, and any numerical model must be able to reproduce the main characteristics of this outcome to assess its validity.

In summary, the main problem to solve can be stated as follows: given the sizes and compositions of both a projectile and a target colliding at a given velocity, what are the resulting size, shape, ejection and rotational velocity distributions of the fragments? All the studies that have been conducted on this problem have tried to provide an answer to this question. The main results obtained by recent collisional models focus on the size and velocity distributions, and we will restrict our discussion on these properties in this paper. The determination of fragments' shape and spin properties requires a higher degree of sophistication that has not been achieved by numerical models yet.

This lecture presents the main principles and results of laboratory experiments and numerical simulations that have contributed to our current understanding of the collisional process in the high impact energy regime, in which the target is disrupted. Scaling laws as well as crater forming events which do not strictly involve the same physical process are not addressed in this lecture. Detailed reviews on these topics can be found elsewhere (see e.g. [21, 22, 31]). Section 2 starts with a brief review of the different methods used to perform laboratory experiments and their results on centimeter-scale targets. Section 3

exposes the theoretical basis of the fragmentation process and the main principles of the model of brittle failure. The numerical methods used to simulate a collision, including both the fragmentation and gravitational phases, are presented in Sect. 4. Section 5 gives the most important results that have been obtained so far by those numerical codes. Although major advances have been performed thanks to those results, the problem is still far to be understood in all its aspects and will certainly remain a subject of long term and intensive research activities. Conclusions and perspectives are presented in Sect. 6.

2 Laboratory Experiments

Laboratory experiments have been the natural first step for studying the collisional process between small bodies. Several methods have been developed to perform impact experiments at speeds that are relevant to the asteroid problem, i.e. up to several kilometers per second. Table 1 summarizes the different kinds of experiments with their range of application and the corresponding references.

The first experimental facility to be used in this purpose was the Vertical Gas Gun Range of NASA's Ames Research Center developed by D.E. Gaults in the early 1960's. Other gas guns have then been developed, for instance, at Kyoto, Japan, by A. Fujiwara, at CalTech in Pasadena, USA, by T. Ahrens and at Johnson Space Center by F. Hörz. A great number of experiments have been performed using these facilities, allowing the development of a large database of outcome properties. This database has been enriched by two other kinds of facilities. Open-air explosive disruption experiments have been carried out in a traversine quarry in Tuscani, led by G. Martelli and I. Gibblin. These explosive experiments have many advantages such as the possibility to achieve higher fragmentation energies than by high velocity impacts. The disruption of larger targets can thus be performed. Moreover, being in an open-air environment these experiments do not suffer from confining boundary conditions such as walls which can cause secondary fragmentation, making difficult the determination of the original fragment size distribution. The reader interested in a review on impact experiments can refer to [28]. The last kind of facility is the high pressure chamber constructed at Boeing Company by K. Housen and R. Schmidt, which can be pressurized to 40 MPa and thus allows to simulate the effect of a confining overpressure on the fragmentation process, which is supposed to happen in the interior of a large asteroid. Rigorously, such overpressure is not exactly the equivalent of the lithostatic pressure inside an asteroid, as the latter depends on the position of the material within the body whereas the overpressure is applied uniformly on the target.

Although, they are limited to small-scale targets and impact energies, the outcomes of these experiments contain some qualitative characteristics which

are still valid at larger scales and which can be looked for with numerical simulations.

Table 1. Experimental impacting techniques, based on [28]

Accelerating technique	Max vel. (m s ⁻¹)	Typical operating velocities (m s ⁻¹)	Ref.
Drop method	50	—	[20] [39]
Pressurised air gun	500	200	[15]
Single stage gas gun	1500	1000	[13]
Powder gun	2000	1000	[38] [30]
Two-stage light gas gun	7000	3000–5000	[48]
Contact charge	6000–8000	7000	[23] [27] [18]
Electromagnetic rail gun	>10000	6000–7000	[25]

2.1 Degree of Fragmentation

By definition, the degree of fragmentation is usually represented by the mass ratio of the largest remnant to the target, called f_l . Obviously, it depends on several factors, such as the impact energy, the internal structure and composition of the bodies that are involved etc. One of the big challenges of collisional studies is to characterize the impact energy required to obtain a given degree of fragmentation as a function of the other factors limited at first to the target's size and projectile's velocity.

By convention, the boundary between the cratering regime and the disruptive one is defined by $f_l = 0.5$, which corresponds to a largest fragment containing 50% of the original mass. Researchers developing scaling laws try to determine the value of the impact energy required to obtain this value for any target's size. Smaller values of f_l correspond to disruptive collisions, whereas higher values correspond to cratering impacts. This is a crude separation since in reality, the cratering regime rather corresponds to higher values of f_l (toward 0.8). However, a distinction remains useful because the cratering regime does not involve exactly the same physical process as the disruptive one. Cratering events are studied by considering a semi-finite body, as the process involves only a limited region of the target, whereas a disruption involves the whole body's structure. In particular, during a disruptive event, the shock wave reaches the side opposite to the impact point of the target and then reflects as a tensile wave. As this lecture concerns only this regime,

the reader can refer to the book of Melosh ([31]) for a good introduction of cratering mechanics.

A wide variety of experiments have been performed at high impact energy, involving several kinds of target material and structure, as well as different impact speeds and geometries. Experiments using similar initial conditions can lead to results that differ quantitatively, due to very small-scale differences and subtle effects, but some global properties still hold true for all of them. In particular, the value of f_l is clearly dependent on both the target's material and the impact energy Q , defined as the projectile's kinetic energy normalized to the mass of the target. For different kinds of material (ice, basalt, etc.), there is only little overlap in the value of Q resulting in a given f_l for different target materials. It is even possible to discriminate between different kinds of materials when only Q and f_l are known. Moreover, for a same material, the degree of fragmentation seems to depend on the impact energy in a way that can be represented by a power-law relationship of the form: $f_l = KQ^{-\alpha}$ where K is a parameter that contains information on the target material and strength while the exponent α is determined to fit the data. Experiments with ice and silicate obtain approximately the same value of this exponent ($\alpha \approx 3/4$) and a higher value is obtained with targets in iron ($\alpha \approx 3/2$), although a much greater number of data would be necessary to determine robust conclusions. Apart from those uncertainties, such a power-law relationship is practical as it can easily be inserted in collisional evolution models of small body populations. However, the value of K and α remain badly constrained and are necessarily limited to impact energies and sizes adapted to these experiments.

2.2 Fragment Size Distribution

The number of fragments generated by the disruption of a target is always found to be inversely proportional to their size in a way that one usually fits using one or several power-laws. The cumulative size distribution of fragments is then represented by the expression: $N(> D) \propto D^{-a}$, where D is the fragment's diameter and $N(> D)$ is the number of fragments with diameter greater than D . The value of the exponent a is badly constrained and no general rule has ever been found that gives a systematic value as a function of the impact parameters. Some experiments find larger values of a with decreasing value of f_l , and other experiments suggest that a two-slope model gives a better representation. Nevertheless, one always finds that the number of fragments increases quite rapidly with decreasing size, and fitting this trend by a power law remains very practical for collisional evolution models of small body populations.

2.3 Fragment Velocity Distribution

Ejection velocities of fragments are very important to determine since they indicate how far a group of fragments extends in space, which has many

consequences in asteroidal studies. However, experiments are confronted to great difficulties for measuring the velocity of individual fragments, and data are always limited to the largest fragments which are the easiest to identify and follow. The method of measurement consists generally in taking high-speed films of the disruption in two views that are orthogonal in order to obtain a velocity vector for each fragment. However, recovering the identity of an individual fragment from those views is not an easy task since the disruption is characterized by a rather thick cloud of dust and small fragments. Therefore, only large fragments can be identified and no robust rule can be derived from such a few data. However, qualitatively, the experiments tend to indicate that larger fragments have lower velocities, and these velocities increase with the impact energy Q .

Ideally, one would like to obtain a rule that links the fragment's size and its velocity. In particular, the existence of a size-velocity relationship would be very well adapted to collisional evolution models. However, given the difficulty of measuring those data, no robust rule has been obtained on the basis of experiments. Nakamura and Fujiwara ([40]) found a certain degree of correlation between the velocity V and the mass m of a fragment that they expressed by: $V(m) = V_0(m/M)^{-r}$, where M is the target's mass, V_0 is a parameter determined by the impact energy Q , and r is chosen to fit the data. But such a relationship does not necessarily represent well data of all experiments, so that its validity is still a matter of debate.

3 Fragmentation Phase: Theoretical Basis

As we have seen in the previous section, laboratory experiments involve targets that are several tens of cm in size at most, while asteroids range up to 1000 km in diameter, which is a factor of $10^6 - 10^7$ bigger. Moreover, outcome properties of experiments, such as size and velocity distributions of fragments are rather scattered, and show only qualitative trends. Therefore, the extrapolation of laboratory results to problems involving asteroids is confronted to major difficulties. Several scaling laws have been developed in this aim, but depending on the assumptions that they use, they can differ by orders of magnitude with respect to each other.

The only way to determine the collisional outcome at asteroid scales without relying on scaling laws is to simulate the process in a direct way. This requires the development of numerical algorithms based on our understanding of the physics involved in the process. This is a great challenge as this physics is poorly known and we must rely on our limited knowledge of the fragmentation process in terrestrial materials. Moreover, the development of numerical algorithms that solve accurately and efficiently the equations that represent this process is not an easy task. However, in the last decade, major advances have been reached in simulations of collisions. In the following I first present the main principles of the fragmentation process as we understand it.

3.1 Basic Equations

Three kinds of waves (elastic, plastic and shock) propagate through a rock during an impact. Elastic waves are well known and determined by linear relationships between the stress and strain tensors. Plastic waves begin to develop when the material strength changes with the wave amplitude. At wave amplitudes that are high enough and associated to shock waves, the material strength is totally reduced, so that it can be ignored. The body is in this case treated as a fluid. Being non-linear, the transitory behaviors between these kinds of waves are difficult to determine analytically from constitutive models. Therefore, to study these phenomena, the best strategy is to implement the bulk properties of a given rock in a numerical model of continuous medium (usually called *hydrocode*), including a yielding criterion and an equation of state for the appropriate material. In this way, no assumption on the form of the stress wave that drives the fragmentation is necessary since the initial conditions evolve numerically based on a rheological model and a failure criterion. The appropriate regime (elastic, plastic or shock) is determined by the computation.

A good description of the propagation of elastic, plastic and shock waves is presented in [31]. The basic equations that must be solved to compute the process are the well-known conservation equations of hydrodynamics that can be found in standard textbooks. The first equation represents the mass conservation. Its expression is:

$$\frac{d\rho}{dt} + \rho \frac{\partial}{\partial x_\alpha} v_\alpha = 0 \quad (1)$$

where d/dt is the lagrangian time derivative. Other variables have their usual meaning and the usual summation rule over repeated indices is used. The second equation describes the momentum conservation (in absence of gravity):

$$\frac{dv_\alpha}{dt} = \frac{1}{\rho} \frac{\partial}{\partial x_\beta} \sigma_{\alpha\beta} \quad (2)$$

where $\sigma_{\alpha\beta}$ is the stress tensor given by:

$$\sigma_{\alpha\beta} = -P\delta_{\alpha\beta} + S_{\alpha\beta} \quad (3)$$

where P is the isotropic pressure and $S_{\alpha\beta}$ is the traceless deviatoric stress tensor. The energy conservation is then expressed by the equation:

$$\frac{du}{dt} = -\frac{P}{\rho} \frac{\partial}{\partial x_\alpha} v_\alpha + \frac{1}{\rho} S_{\alpha\beta} \dot{\epsilon}_{\alpha\beta} \quad (4)$$

where $\dot{\epsilon}_{\alpha\beta}$ is the strain rate tensor given by:

$$\dot{\epsilon}_{\alpha\beta} = \frac{1}{2} \left(\frac{\partial}{\partial x_\beta} v_\alpha + \frac{\partial}{\partial x_\alpha} v_\beta \right) \quad (5)$$

This set of equations is still insufficient in the case of a solid since the evolution in time of $S_{\alpha\beta}$ must be specified. The basic Hooke's law model is assumed in which the stress deviator rate is proportional to the strain rate:

$$\frac{dS_{\alpha\beta}}{dt} = 2\mu \left(\dot{\varepsilon}_{\alpha\beta} - \frac{1}{3}\delta_{\alpha\beta} \right) + S_{\alpha\gamma}R_{\beta\gamma} + S_{\beta\gamma}R_{\alpha\gamma} \quad (6)$$

where μ is the shear modulus and $R_{\alpha\beta}$ is the rotation rate tensor given by:

$$R_{\alpha\beta} = \frac{1}{2} \left(\frac{\partial}{\partial x_\beta} v_\alpha - \frac{\partial}{\partial x_\alpha} v_\beta \right) \quad (7)$$

This term allows the transformation of the stresses from the reference frame associated with the material to the laboratory reference frame in which the other equations are specified.

This set of equations can now be solved, provided an equation of state is specified, $P = P(\rho, u)$, linking the pressure P to the density ρ and internal energy u . The Tillotson equation of state for solid material ([51]) is generally used. Its expression and method of computation, as well as parameters for a wide variety of rocks are described by Melosh ([31], Appendix II). Other equations of states have been developed such as ANEOS ([50]); all have different pros and cons and remain necessarily limited to materials studied in laboratory. This is one of the limits of any collisional model that necessarily relies on the behavior of known materials that do not necessarily represent the materials constituting an asteroid.

Perfectly elastic materials are well described by these equations. Plastic behavior beyond the Hugoniot elastic limit is introduced in these relations by using the von Mises yielding criterion. This criterion limits the deviatoric stress tensor to:

$$S_{\alpha\beta} = f S_{\alpha\beta} \quad (8)$$

where f is computed from:

$$f = \min \left[\frac{Y_0^2}{3J_2}, 1 \right], \quad (9)$$

where J_2 is the second invariant of the deviatoric stress tensor, defined as:

$$J_2 = \frac{1}{2} S_{\alpha\beta} S_{\alpha\beta} \quad (10)$$

and Y_0 is a material dependent yielding stress which generally depends on temperature, density, etc. in such a way that it decreases with increasing temperature until it vanishes beyond the melting point.

3.2 Fundamental Basis of Dynamical Fracture

The von Mises criterion is adapted to describe the failure of ductile media such as metals. Brittle materials like rocks do not undergo a plastic failure but

rather “break” if the applied stresses exceed a given threshold. Conversely, the yielding beyond the Hugoniot elastic limit does not prescribe any permanent change in the constitution of the material, since once stresses are reduced the original material remains behind, possibly heated by the motion against the remaining stress, but otherwise not weakened. Therefore, it is not adapted to impacts into rocks, as any yielding beyond the elastic limit invariably involves irreversible damage, and one needs to know how the rock is permanently altered by the event. A realistic fracture model is then clearly required to study the disruption of a solid body. In the following, I present the main principles of the Grady-Kipp fragmentation theory ([19]), which is at the basis of fracture models implemented in numerical simulations of asteroid disruption.

The Grady-Kipp theory is based on the nucleation and propagation of flaws that are initially pre-existing in a rock. The original continuum model of fragmentation of Grady-Kipp is statistical and has been shown to lead to inconsistencies between hydrodynamics and fragmentation unless the flaws are made explicit ([6]). However, the bulk of the theory remains valid and all the issues related to technical or computational implementation are beyond the scope of this lecture (see [6] for their description).

The brittle failure is characterized by a sudden stress release when the material is loaded to a limit called its strength. For a linear elastic solid, before reaching this limit, the stress increases linearly with the strain. While the load is applied, small microcracks nucleate and propagate through the rock. Crack tips seek minimal-energy paths through the undamaged rock in response to the increasing stress. Then, communication between flaws can occur as a result of the stress release due to crack propagation, which can eventually lead to catastrophic rupture. Brittle failure thus relies on the assumption of the existence of incipient flaws in the rock. Their presence has been observed in natural rocks and is usually described by the so-called Weibull distribution.

The Weibull Distribution of Incipient Flaws

A two-parameter Weibull distribution describes the network of incipient flaws in any material:

$$N(\varepsilon) = k\varepsilon^m \quad (11)$$

where ε is the strain and N is the number density of flaws that activate at or below this value of strain. The Weibull parameters m and k are material constants which have been measured for a number of geological and industrial materials, although data are quite scarce for some important rocks (see [32]). In particular the parameter k varies widely between various rock types and the exponent m ranges typically between 6 and 12. Their values represent important material properties as large values of m describe homogeneous rocks with uniform fracture threshold, while small values apply to rocks with widely varying flaw activation thresholds. The existence of incipient flaws within any

rock is understood to originate from its cooling history and from crystal lattice imperfections. Due to the initial presence of these flaws, when a finite strain rate $\dot{\epsilon}$ is applied, a stress increase occurs in time, which is moderated by the propagation of active flaws. Thus a competition takes place between the stress increase due to loading and the stress release due to flaw activation and propagation, until a temporary equilibrium is reached at the time of peak stress. Then, the stress decreases to zero as active flaws propagate rapidly through the rock.

From these explanations, it is obvious that the crack growth velocity c_g is an important parameter since it governs the stress release due to an active flaw. Experiments indicate that it relates to the speed c_l of longitudinal waves in a rock by $c_g \approx 0.4 \times c_l$. Under moderate conditions, since cracks propagate at this fixed velocity, the weakest flaws (those which activate at lower values of ϵ) suffice to accommodate the growing stresses. Therefore, the peak stress at failure is low and fragments are relatively large (see Fig. 1). Conversely, more resistant flaws have time to activate at high strain rates. In this case, the peak failure stress is high and fragments are small. This process depends strongly on the adopted value of the crack growth velocity c_g . In particular, fragment sizes scale with c_g . For instance, a higher velocity would enhance the efficiency at which a crack relieves stress, since stress release is proportional to crack length cubed. As a consequence, fewer flaws would be required to relieve a given increase in stress.

Consequences on the Definition of the Material Strength

From the previous explanations, one may conclude that defining the material strength as *the stress at which sudden failure occurs* is not rigorously adequate. Material strength could rather be defined as the stress at which the first flaw begins to fail, thereby initiating an inelastic behavior characterized by irreversible deformation. But in practice, the adopted definition is the peak stress which the rock undergoes prior to failure. It is then not a material constant, since as explained above, the peak stress is a function of the loading history of the rock. This is the reason why a distinction is made between static strength and dynamic strength on the basis of the loading rate. For extremely small loading rates, elastic stresses increase in equilibrium until the onset of catastrophic failure. This occurs at loading rates that are typically smaller than $\approx 10^{-6}$ strains per second. Static tensile strength decreases with increasing size of the rock due to the greater probability of finding a weaker (larger) flaw. At high enough loading rates, stresses can continue to build while catastrophic rupture has begun. In this case, it is more appropriate to speak of dynamic failure. For most rocks, dynamic strain rates are of the order of 1 s^{-1} and decrease with increasing rock size. Therefore, the peak stress that the rock suffers prior to failure is rigorously called the material's dynamic strength at that strain rate. This dynamic strength increases with strain rate and is always greater than the static strength. All hypervelocity impacts into

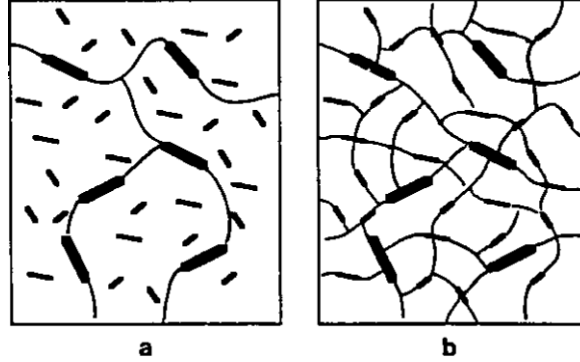


Fig. 1. Dynamic fragmentation illustrated using low and high strain rates. Starting with a Weibull distribution of flaws (shown as lines, with the weakest being the longest), cracks propagate by the activation of these flaws according to their threshold strains. Stresses are then relieved and cracks coalesce to form fragments. **(a)** At low strain rates, the strongest and by far most numerous flaws are never activated because the weak flaws have enough time to relieve the stresses. The resulting fragments are thus large and the peak stress of failure is low; **(b)** at high strain rate, since the rate of crack growth (c_g) is finite, the weak flaws are not numerous enough to relieve the growing stresses and strong flaws can then be activated, leading to smaller fragment sizes and a higher failure stress. A relationship between dynamic tensile stress and strain rate has been derived from these properties, $\sigma \propto \dot{\epsilon}^\gamma$ ($\gamma \approx 1/4$), and fragment size L and strain rate s are also linked by the relation $L \propto 1/s$. Kindly provided by E. Asphaug ([1])

small targets are in the dynamic regime, but some impacts on large bodies can still be in a regime close to the static one. In this case, close to the impact point, part of the event can be dynamic, but some important aspects can also be understood in terms of quasi-static failure.

Finally, note that two kinds of strength can also be considered. Depending on the nature of the loading, one can speak of compressive strength or tensile strength. For a given material, the tensile strength is much lower than the compressive one. However, brittle failure occurs only in response to tensile and deviatoric stresses, even in the case of a compressive applied loading axis. One limitation to the study of asteroid disruption concerns the adopted values of tensile strength in models since these values necessarily correspond to materials for which published tensile strengths are available at loading rates from quasi-static to intensely dynamic, and which may not represent well asteroid materials.

Damage Propagation

When a flaw becomes active, i.e. once the local tensile strain ϵ has reached the flaw's activation threshold, a crack begins to grow at velocity c_g and this

growth affects the dynamics of the system by releasing local stresses. To compute this local stress-release, Grady and Kipp introduce a state variable D (for “Damage”) which ranges from 0 to 1 and regulates the elastic modulus when the material is in tension. In other words, damage expresses the reduction in strength under tensile loading so that the damage-relieved stress becomes:

$$\sigma_D = \sigma(1 - D) \quad (12)$$

where σ is the elastic stress in the absence of damage. When $D = 0$, the material is intact and unfractured, whereas when $D = 1$, the material is completely fragmented. It then corresponds to a collection of unconnected fragments or a cohesionless fluid, and offers no resistance to further extensional strength. A material with sufficient cracks to have $D = 0.5$ feels half the stress it would otherwise feel under the same tensile strain.

Walsh has shown that a crack relieves stresses in a volume approximately equal to its circumscribing sphere ([52]). Therefore, to relate the growth of cracks to damage variable D , this variable is defined locally as the fractional volume that is relieved of stress by local growing cracks:

$$D = \frac{4}{3} \frac{\pi}{V} a^3 \quad (13)$$

where $V = 4\pi R_s^3/3$ is the volume in which a crack of half-length R_s is growing, and the half-length of a growing crack at time t is $a = c_g(t - t_0)$, where t_0 is the crack activation time. From these expressions, a convenient differential equation is derived for the time evolution of the damage variable assigned to a sphere of radius R_s in which a crack is growing:

$$\frac{dD^{1/3}}{dt} = \frac{c_g}{R_s} . \quad (14)$$

The overall damage experienced by the material is then given by the distribution of D over the entire object. In an elemental subvolume, once the activation threshold has been exceeded, damage is independent of strain, whereas over a larger volume it depends on it, since the number of active flaws is strain-dependent. The procedure above would then give accurate results only in the limit of infinite resolution, as all flaws could then be computed explicitly. Grady and Kipp take instead a purely statistical approach and compute in the Weibull distribution the number of flaws that have activated, as a function of time, for a given loading rate, and integrate the stress-relieved volume due to all the flaws that have grown since the first one activated ([19]). Their resulting integral equation reduces to the following differential expression in the limit of constant strain rate:

$$\frac{dD^{1/3}}{dt} = \frac{m+3}{3} \alpha^{1/3} \varepsilon^{m/3} . \quad (15)$$

In this expression α is a material constant that depends on k , m and c_g . A 2D hydrocode that uses this damage expression has first been developed ([32]). Later, Benz and Asphaug have found a way to combine the two

approaches by modelling explicit flaws, using the expression previously indicated of the time evolution of damage assigned to a sphere in which a crack is growing, together with the statistical flaws that are required due to resolution constraints in order to complete the distribution. In this way, their model, which details can be found in their original article ([6]), reduces to the Grady-Kipp approach in the limit of coarse computational resolution, and becomes purely explicit at the other extreme.

3.3 Numerically Simulating the Fragmentation Phase

To develop a numerical algorithm that solves the equations described in previous sections, it is necessary to transform this set of equations, which represents a space-time continuum, into algebraic expressions which are the only ones that a computer can solve. Such a discretization necessarily modifies the original equations by introducing error terms and the challenge of numerical hydrodynamics is to find algebraic expressions that minimize these errors.

To perform this transformation, it is first necessary to define the coordinate system in which the equations have to be solved. Some hydrocodes use an Eulerian frame, which corresponds to a fixed coordinate system (or moving at a prescribed velocity), but the most appropriate approach is to use a Lagrangian frame which is attached to and moves with the material. Indeed, recall that fragmentation is the result of the entire stress history experienced by a given piece of material. Therefore, a Lagrangian approach in which the frame of references is attached to the material is the only really suitable framework for solving these equations. However, if a grid is used to define the asteroid, there are great limitations on the extent of deformation achievable with multi-dimensional Lagrangian codes based on a classical finite difference scheme, since the numerical derivatives become less and less accurate if not totally wrong. This is the reason why simulations using grid-methods usually stop very early in the evolution of the impact process as they are unable to follow the large distortion associated to late times.

The 3D Lagrangian hydrocode developed by Benz and Asphaug ([6, 7]) represents the state-of-the-art in numerical computations of dynamical fracture of brittle solids. It does not suffer from the problems specified above as it uses the method called Smooth Particle Hydrodynamics (SPH). Complete details about this method can be found in a review by Benz ([4]). Basically, the value of the different hydrodynamics quantities are known at a finite number of points which move with the flow. Starting from a spatial distribution of these points called particles, the SPH technique allows to compute the spatial derivatives without the necessity of an underlying grid. Once these derivatives have been computed and forces determined, the system is integrated over time in usual ways. Thus, this technique allows to avoid the use of a grid and differs widely from grid-based methods not only conceptually but also because of the numerical requirements and algorithms that allow a good efficiency. All details can be found in the already mentioned papers by Benz

and Asphaug. Using this technique, the 3D SPH hydrocode is then able to simulate consistently from statistical and hydrodynamical points of view the fragments that are smaller or larger than the chosen resolution (number of SPH particles). Statistical flaws (microscopic) are propagated at the particle scale according to the Grady-Kipp fracture model, while actual flaws (macroscopic) are solved in a manner that does not depend on the resolution. The method thus guarantees that any increase in resolution does not alter the physics of fracture, but only modifies the accuracy. The resulting system has proven to predict the sizes, positions and velocities of the largest fragments of laboratory experiments with a precision never achieved before.

3.4 Summary of Limitations Due to Material Uncertainties

Besides the difficulty to develop a numerical algorithm that solves consistently the appropriate set of differential equations, one of the great limitation of any research on collisions between small bodies in the Solar System comes from the uncertainties on material properties of the objects involved in the event. For instance, ten material parameters describe the usually adopted Tillotson equation of state. Other sensitive material-dependent parameters are the shear modulus, two parameters describing the fracture model, and the value of the crack growth velocity, usually taken to be 40% of the longitudinal sound speed and which high influence has been emphasized in previous sections. Laboratory experiments have provided direct measurements of many of those parameters for ordinary material. However, even though the 3D SPH hydrocode has been able to reproduce successfully experiments using appropriate parameters, nothing tells us that these parameters apply to asteroid or cometary materials, since those are essentially unknown. Therefore, one has to keep in mind that the ability to compute the fragmentation of a target with known material properties is a necessary check to prove the reliability of the method, but the outcomes may well be different in the case of drastically different asteroid material compositions. Until a sample return mission is achieved, allowing to obtain material parameters from at least one real asteroid (for instance, the Japanese mission Hayabusa launched in May 2003 should bring back a sample of a 500-meter size Near-Earth Object named Itokawa in 2007), our simulations of asteroid disruption will necessarily rely on ordinary material parameters. However, despite this limitation, as I will show further, recent studies of asteroid family formation based on numerical simulations have been able to reproduce the main properties of real families formed in several impact energy regimes.

4 Gravitational Phase: Large-Scale Simulations

The fragmentation of an asteroid larger than tens of kilometers can generate several 10^5 kilometer-size fragments. The early stages of the gravitational

evolution of these fragments is still part of the collisional process. Current computer performances do not allow to simulate this phase on reasonable times by using traditional N-body codes based on direct force calculations. Indeed, CPU time required to compute the evolution of N particles with such codes scales with N^2 .

The last results that have been obtained on the collisional process at asteroid scale have taken advantage of the creation of new and fast methods to compute N-body interactions. More precisely, a modified version of a cosmological N -body code, called *pkdgrav*, which is based on the so-called hierarchical tree method has been used to compute the gravitational interactions (see e.g. [45]). This code can be run on both shared-memory and distributed-memory parallel architectures. The tree component of the code provides a convenient means of consolidating forces exerted by distant particles, reducing the computational cost. The parallel component divides the work evenly among available processors, adjusting the load each timestep according to the amount of work done in the previous force calculation. The code uses a straightforward second-order leapfrog scheme for the integration and computes gravity moments from tree cells to hexadecapole order.

An important aspect of the gravitational phase is that some fragments may reaccumulate with each other, depending on their relative velocities, to form larger bodies. Therefore, it is fundamental to account for this aspect in the computation. A big advantage of *pkdgrav* is that it treats collisions and mergers between particles. Indeed, particles are considered to be finite-sized hard spheres and collisions are identified each step using a fast neighbor-search algorithm. Whenever two fragments collide they can merge creating a new fragment located at and moving with the center of mass of the system and with a mass equal to the sum of the individual masses. The radius of the new fragment is computed from knowing the mass and the density and assuming spherical shape. A merging criterion based on relative velocity and angular momentum is defined such that fragments are allowed to merge only if their relative velocity is smaller than their mutual escape velocity and/or when the resulting spin of the merged fragment is smaller than the threshold value for rotational fission. Non-merging collisions are modeled as bounces between hard spheres whose post-collision velocities are determined by the amount of dissipation taking place during the collisions. The latter is determined by the coefficient of restitution in the tangential and normal directions (see [44], for details on this computation). Since the values of these coefficients are poorly constrained, they are arbitrarily set equal to 0.5 in the results presented further.

Despite these great sophistications, the computation of the gravitational phase still contains some simplifications and a great amount of work remains necessary and is under way to increase its degree of realism. In particular, fragment shapes are not accounted for currently, as all particles remain spherical and aggregates formed by reaccumulation are represented by a single spherical particle with corresponding mass.

5 Current Understanding and Latest Results

In the following, the most recent results obtained by simulating numerically the catastrophic disruptions of large asteroids are presented. The simulations rely on the theory of fragmentation of solid bodies exposed in Section 3 and have been performed using the 3D SPH hydrocode and the gravitational N -body code mentioned in previous sections.

Recall that asteroid families represent a unique laboratory that Nature offers us to study large-scale collisional events. Indeed, observed asteroid families in the main asteroid belt are each composed of bodies which originally resulted from the break-up of a large parent body (e.g. [29]).

Interestingly, until recently, the theory of the collisional origin of asteroid families rested entirely on these similarities in dynamical and spectral properties and not on the detailed understanding of the collisional physics itself. Indeed, laboratory experiments on centimeter-scale targets, analytical scaling rules, or even complete numerical simulations of asteroid collisions were not able to reproduce the physical and dynamical properties of asteroid families (e.g. [47]). The extrapolation of laboratory experiments to asteroidal scales yields bodies much too weak to account for both the mass spectrum and the dynamical properties of family members. More precisely, in a collision resulting in a mass distribution of fragments resembling a real family, which can contain many big members, the ejection velocities of individual fragments are much too small for them to overcome their own gravitational attraction. The parent body is merely shattered but not dispersed and therefore no family is created. Conversely, matching individual ejection velocities and deriving the necessary fragment distribution results in a mass spectrum in which no big fragment is present, contrary to what is indicated by most real families (e.g. [12, 14]).

The collisional origin of asteroid families thus implies that not only the parent body (up to several hundreds kilometers in size) has been shattered by the propagation of cracks but also that the fragments generated this way typically escape from the parent and reaccumulate elsewhere in groups in order to build up the most massive family members. Such a process had already been suggested (e.g. [11]), and the possibility that at least the largest fragment from a collision consists in a rubble pile had also been indicated later by means of numerical simulations [8]. The effect of gravitational reaccumulation was then estimated by a procedure which consists in searching for the largest group of gravitationally bound debris immediately following the collision, and not in computing explicitly the gravitational interactions between the fragments. It is only very recently that the formation of many large family members by reaccumulation of smaller fragments has been demonstrated explicitly by Michel and his colleagues [34–37].

Such aggregates of gravitationally bound fragments, as the ones formed in the last simulations, are usually defined as *rubble-piles* in the asteroidal community, which means that they are *loose* aggregates of fragments held

together by gravity. A detailed definition and a review of this topic are presented in [46]. Roughly, such bodies have little to no tensile strength, i.e. they can be torn apart easily by planetary tides. Only indirect evidence for such structures exist. Indeed, the structural properties of asteroids are difficult to establish since directly measurable quantities do not distinguish between solid bodies and rubble piles. Rubble piles have been invoked to explain, for instance, the low density of some observed bodies like 253 Mathilde whose measured density by the *NEAR* probe is 1.35 g/cm^3 ([55]) or the lack of fast rotators among asteroids with sizes larger than a few hundreds meters ([43]).

For the first time, Michel and his colleagues have simulated entirely and successfully the formation of asteroid families in two extreme regimes of impact energy leading to either a small or a large mass ratio of the largest remnant to the parent body M_{lr}/M_{pb} ([34]). Two well-identified family have been used for comparison with simulations: the Eunomia family, with a 284 km parent body and $M_{lr}/M_{pb} = 0.67$, has been used to represent the barely disruptive regime, whereas the Koronis family, with a 119 km parent body and $M_{lr}/M_{pb} = 0.04$, represented the highly catastrophic one. In these simulations, the collisional process was carried out to late times (typically several days), during which the gravitational interactions between the fragments could eventually lead to the formation of aggregates or rubble piles far from the largest remnant. It was first assumed somewhat unrealistically that particles colliding during the gravitational phase always stuck perfectly and merged regardless of relative velocity and mass. Michel and his colleagues then improved on this treatment by allowing for the dissipation of kinetic energy in such collisions and applying an energy based merging criterion ([35]). This improved treatment did not change the conclusion obtained with the more simplistic one, and the reaccumulation process remains at the origin of large family members.

5.1 Disruption of Monolithic Family Parent Bodies

The two studies mentioned above used as a starting condition a monolithic parent body represented by a basalt sphere. The projectile's parameters (diameter, velocity and impact angle) were defined such that the expected value of M_{lr}/M_{pb} was successfully obtained by the simulation (see [34, 35] for details). A first unexpected result was found in all the explored impact energy regimes: the fragmentation phase always leads to the complete pulverization of the parent body down to a fragment size corresponding to the resolution limit (of the order of 1 km). Then, the gravitational phase during which these fragments interact leads to many reaccumulations (see Fig. 2). Eventually, the fragment size distribution is dominated by aggregates formed by reaccumulation of smaller fragments and only the smallest size end of the distribution consists in individual (or *intact*) fragments. Moreover, for each family, the simulated distribution is qualitatively compatible with that of real family members, even though some large fragments always lack between the largest

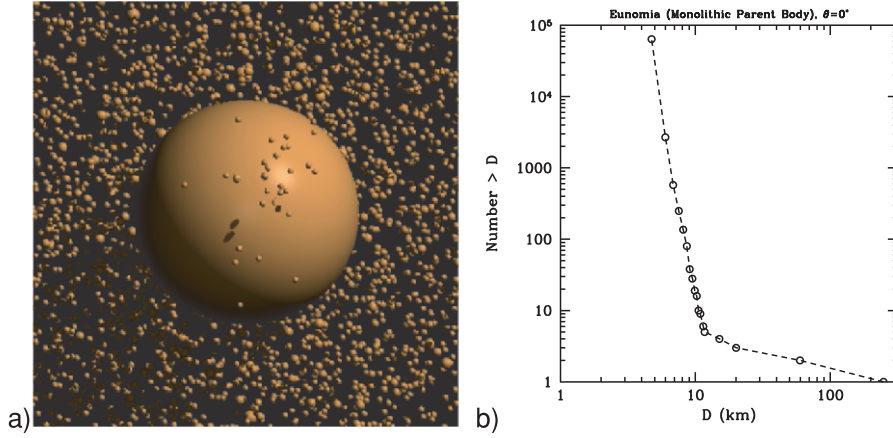


Fig. 2. (a) Image taken from the simulation of the gravitational phase of the Eunomia family formation ≈ 84 minutes after fragmentation of the parent body, when the largest fragment of Eunomia is almost already formed; it is represented by the large sphere at the center of the image which is produced by the reaccumulations of smaller ones; (b) Cumulative diameter distribution (in km) of the fragments resulting from the simulation of the Eunomia family formation in a log-log plot; the simulation was performed using a target with diameter and bulk density corresponding to the parent body of the family. The number of particles used to define the target was set to 2×10^5 . The impact angle was $\theta = 0^\circ$ and the impact velocity of the projectile was 5 km s^{-1} . These results are described in the paper of Michel and his colleagues ([35])

and smaller ones, compared to the observed distribution (this difference will disappear using pre-shattered parent bodies; see further). Another important result is that many binary systems are formed during the gravitational phase, so that the formation of asteroid satellites can be seen as a natural outcome of catastrophic collisions. However, the timescale of the simulations is too short and external perturbations (solar tides, etc...) should be included to determine their long-term stability and lifetime. Nevertheless, their occurrence provides a possible explanation of the origin of some of the observed ones. The interested reader can find a detailed study of satellite formation during collisions in a recent paper ([17]).

As for the orbital dispersion of fragments obtained from these simulations, it is always smaller than the dispersion of real family members computed from their proper orbital elements ([24,56]). However, the computation of the dispersion of simulated fragments is confronted to many unknowns such as the values of the true anomaly and perihelion argument of the parent body at the instant of impact, which are required by Gauss' formulae to convert ejection velocities into orbital elements (see e.g. [35]). Nevertheless, the compact orbital

dispersions found in simulations are still consistent with the larger ones of real families. Indeed, it has been realized that the current orbital dispersion of family members may not represent the original one following the break-up of the parent body. The post-collisional evolution of family members is affected by dynamical mechanisms such as high order resonances and/or the thermal Yarkovsky effect which can both cause a diffusion of the orbital elements over timescales long enough (see e.g. [9, 10, 42]). When the estimated age of the considered family and the diffusion timescale of the orbital elements of family members due to these mechanisms are taken into account, the comparison between the compact dispersions obtained by the simulations and the larger one of real members leads to a good agreement.

5.2 Disruption of Pre-Shattered Parent Bodies

Starting from monolithic family parent bodies already allowed to reproduce qualitatively the main properties of real family members. However, the cumulative size distribution of simulated fragments is quite systematically characterized by a lack of intermediate sized bodies and a very steep slope for the smaller ones (see plot b) of Fig. 2). Such characteristics are not observed in the size distribution of real family members, which generally looks rather continuous. It is then important to determine whether this systematic effect is a general failure of the simulations, or whether it depends on the initial parameters and assumptions used to start the simulations. In particular, instead of being monolithic, the parent bodies may be modelled with an internal structure composed of different zones of voids and fractures, as if they had first been pre-shattered during their collisional history before undergoing a major event leading to their disruption. If a better agreement could be obtained by playing on the original internal structure of the parent body, this would be an important result with many implications. For instance, simulations could then be used to constrain the internal structure of parent bodies at the origin of known asteroid families that allows to produce the corresponding member properties.

The assumption that large parent bodies are pre-shattered before being disrupted is appropriate not just because it may potentially lead to a closer match with observed properties. The assumed pre-shattered state is thought to be a natural consequence of the collisional evolution of main belt asteroids. Indeed, several studies have indicated that for any asteroid, collisions at high impact energies leading to a disruption occur with a smaller frequency than collisions at lower impact energies leading to shattering effects only (see, e.g. [3, 16, 46]). Thus, in general, a typical asteroid gets battered over time until a major collision eventually disrupts it into smaller dispersed pieces ([33]). Consequently, since the formation of an asteroid family corresponds to the ultimate disruptive event of a large object, it is reasonable to think that the internal structure of this body has been modified from its primordial state by

all the smaller collisional events that it has suffered over its lifetime in the belt.

The size and velocity distributions of family members are the major constraints that can be compared with results of simulations starting from different models of parent bodies. However, as already stated, these distributions can be modified over time by collisional erosion and dynamical diffusion, depending on the proximity of diffusion mechanisms to family members and on the family's age. Recently, a very young family, called Karin, has been identified thanks to the increasing database of asteroid proper elements. Its young age of 5 Myr has been estimated by integrating numerically the orbits of its members backward in time from their current state to the state where the heliocentric orbits of all cluster members and their orientations were nearly the same ([41]). The authors estimated that this convergence could not happen by chance. The Karin family is thus so young that it provides a unique opportunity to study a collisional outcome almost unaffected by erosion and dynamical diffusion.

Michel and his colleagues have recently reported on numerical simulations aimed at determining which classes of collisions reproduce the main Karin family characteristics ([36]). In order to study how the collisional outcome depends on the internal structure of the parent body, they considered different models of such a body and searched for the one which, once disrupted, best matched the observed properties of the Karin family. More precisely, two kinds of parent bodies, both spherical in shape, were considered: (1) purely monolithic, and (2) pre-fragmented (pre-shattered with damage zones but no internal voids – see Fig. 3 – or rubble pile with internal voids). The simulations then showed that the heterogeneities introduced in this way result in significant changes in the size spectrum of the final fragments left after the collision, even though in both cases the targets are completely shattered by the impact. In particular, the authors concluded that the Karin family must have originated from the breakup of a pre-fragmented parent body. Indeed, whereas the fragment size distribution obtained from a monolithic parent body still lacks some fragments at intermediate sizes, disruptions involving a pre-fragmented (pre-shattered or rubble pile) parent body on the other hand result in a much more continuous cumulative size distribution of fragments. Hence, the presence of intermediate sized bodies, as observed in the real family, is most likely a direct consequence of the presence of large-scale fractures or big blocks within the parent body. Since such large members cannot be obtained starting with a monolithic parent body regardless of the impact geometry and material type in all the investigated cases, the authors conclude that the parent body of the Karin family must have been pre-fractured or reaccumulated before suffering the disruption that formed the family. This is consistent with the fact that the Karin parent body actually belonged to the older Koronis family, which, as the authors had shown in previous simulations, is composed of large members that are necessarily reaccumulated objects ([34, 35]).

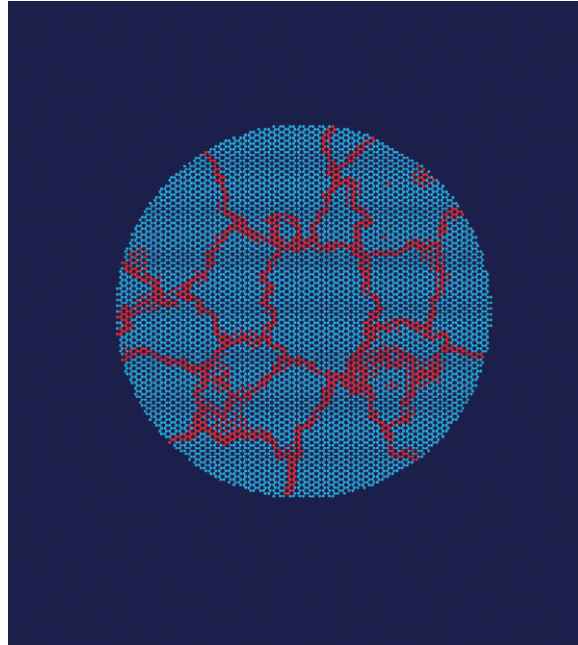


Fig. 3. Slice through the equatorial plane of the 25 km diameter pre-fractured Karin family parent body. The 400 m thick slice shows the network of cracks (in red, defined by totally damaged particles) which divide the parent body into 40 distinct fragments. Note that there are no voids in the parent body and that there is no density discontinuity between the fractured (*red*) and solid (*blue*) material

The authors also remarked that intermediate size fragments are also present in most major asteroid families, implying that many parent bodies in the asteroid belt were probably pre-fractured, and confirmed this interpretation by new simulations ([37]). More precisely, they remade their simulations of the Eunomia and Koronis family formations using pre-shattered parent bodies and compared the results with the ones obtained with a monolithic parent body. The best agreement was again found with pre-shattered parent bodies. In particular, in the case of the Koronis family, an interesting result has been obtained from these simulations, which may have important implications concerning the real family history. The size distribution obtained from the disruption of a pre-shattered parent body contains four largest fragments of approximately the same size (Fig. 4). This peculiar characteristic is shared by the real family, and has been a source of debate as it was assumed that a single collisional event cannot produce such a property (see [37] for a discussion). Moreover, the simulation using a monolithic parent body did not result in such a distribution. The authors then demonstrated numerically for the first time, by using a pre-shattered parent body, that these fragments can actually be produced by the original event, and therefore no subsequent mechanism

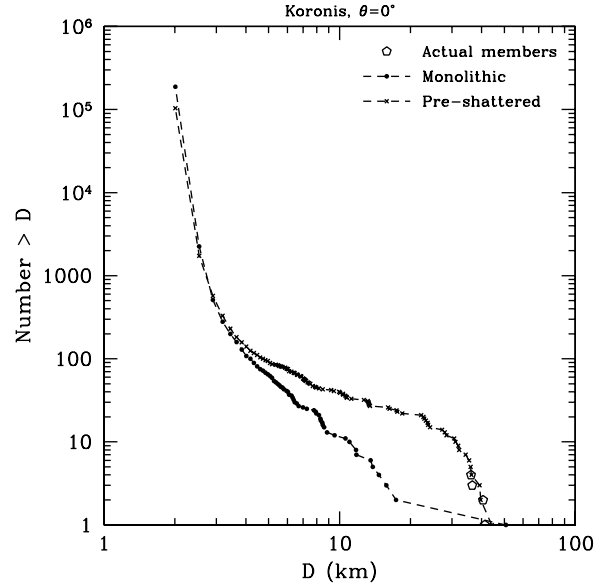


Fig. 4. Cumulative diameter distributions (in km) in a log-log plot for the fragments of the simulated Koronis family ([37]). An impact at 5 km s^{-1} with an angle of incidence θ equal to 0° gave rise to these distributions. Different symbols are used depending on the different parent body models used in the simulations as indicated on the plots. The estimated sizes of the four actual largest members are also shown on the plot ([49]). Simulations of the disruption of a pre-shattered parent body are able to produce such a cluster of four fragments with similar sizes. The smallest fragment (resolution) is limited to 1 km in radius

needs to be invoked to form them, which would require a revision of the entire family history ([37]). According to those results showing that even old families may well have been originated from pre-shattered parent bodies, the authors concluded that most large objects in the present-day asteroid belt may well be pre-shattered.

The second important constraint concerns the ejection velocity distribution of simulated fragments, which must be compatible with that of real family members. Karin's young age ($\approx 5 \text{ Myrs}$) implies that dynamical diffusion has had no time to alter appreciably the initial ejection velocity distribution, which is not the case for the other much older families. Again, they found that a better match is obtained from the pre-shattered parent body. Conversely, the dispersion of the larger fragments produced by a monolithic parent body is definitely too small with respect to that of real members.

In summary, a good match of both the continuous size distribution and the orbital dispersion of the Karin family requires that the parent body was pre-fragmented or reaccumulated. This result supports the overall picture that all large members of asteroid families are reaccumulated bodies, given that

the Karin cluster is at least a second-generation family in the lineage of the older Koronis parent body.

Another big interest of these simulations is that they can be used to determine the impact energy needed to produce a given degree of disruption as a function of the internal structure of the parent body. It is generally found that disrupting a pre-shattered target requires less energy per unit mass than disrupting a monolithic body with the same degree of disruption. This, at first glance a surprising result, is related to the fact that the fractures as modelled by the authors (no porosity and no material discontinuities except damage) do not affect shock waves but only tensile waves. Hence, fragments can be set in motion immediately upon being hit by the shock wave without having to wait for fracture to occur in a following tensile wave. Thus, transfer of momentum is more efficient and disruption facilitated. Note that the presence of large voids such as those in rubble piles affect the propagation of the shock wave, thus reversing this trend. A more detailed study of these properties is in development, but it is already clear that the internal structure of an asteroid plays an important role in the determination of its response to impacts (see also [2]). This is not only relevant for estimating the collisional lifetime of a body in the asteroid belt, but also for developing strategies to deflect a potential Earth impactor.

6 Conclusions

In recent years, thanks to the development of sophisticated numerical codes and the advances in computer technologies, a great step has been achieved in our understanding of the collisional process at large asteroid scales. Results from laboratory experiments on centimeter-size targets cannot be easily extrapolated to those scales but they remain useful to validate at small scales the developed numerical tools. The most recent simulations of asteroid collisional disruptions have successfully reproduced the main characteristics of big asteroid families in several impact energy regimes. They have shown that the collisional process at large scales does not only involve the fragmentation of the parent body, as it is the case in laboratories, but also the gravitational interactions of the fragments. Those interactions can eventually lead to the formation of large aggregates produced by the reaccumulation of smaller fragments. According to these results, most large family members consist in rubble-piles and not in monolithic bodies. Moreover, the internal structure of the parent body itself greatly influences both the impact energy required to achieve a given degree of destruction and the collisional outcome. In particular, pre-shattered parent bodies generate in general a much larger number of big fragments – formed by gravitational reaccumulation of smaller ones – than their monolithic counterpart. The resulting size distributions are then in better agreement with those of real families, suggesting that most of these families originated from pre-shattered parent bodies. This agrees with the

idea that large asteroids get battered over time by small impacts which modify their surface and internal structure until a large event eventually disrupts them.

Although a big step has been accomplished, many uncertainties still remain, even concerning the physical process itself. All numerical codes are limited by the physics one puts into them, which is still based on some models, and by the knowledge of physical parameters that are valid for terrestrial materials. It is however comforting that despite these uncertainties, the results are already consistent with observational constraints provided by asteroid families, but one should keep in mind those uncertainties until physical data directly appropriate to asteroids become available. The future space missions devoted to in-situ measurements and sample returns will certainly help to improve our knowledge.

References

1. E. Asphaug: Dynamic fragmentation in the solar system: applications of fracture mechanics and hydrodynamics to questions of planetary evolution. PhD Thesis, The University of Arizona (1993)
2. E. Asphaug, S.J. Ostro, R.S. Hudson, D.J. Scheeres, W. Benz: *Nature* **393**, 437 (1998)
3. E. Asphaug, E.V. Ryan, M.T. Zuber: Asteroid Interiors. In: *Asteroids III*, ed by W.F. Bottke JR, A. Cellino, P. Paolicchi, R.P. Binzel, T. Gehrels (Univ. of Arizona Press, Tucson 2003), pp 463–484
4. W. Benz: Smooth Particle Hydrodynamics – A Review. In: *Proceedings of the NATO Advanced Research Workshop on The Numerical Modelling of Nonlinear Stellar Pulsations Problems and Prospects*, ed by J. Robert Buchler (Kluwer Academic Publishers, Dordrecht 1990)
5. W. Benz, W.L. Slatery, A.G.W. Cameron: *Icarus* **74**, 516 (1988)
6. W. Benz, E. Asphaug: *Icarus* **107**, 98 (1994)
7. W. Benz, E. Asphaug: *Comput. Phys. Comm.* **87**, 253 (1995)
8. W. Benz, E. Asphaug: *Icarus* **142**, 5 (1999)
9. W.F. Bottke, D. Vokrouhlický, M. Borz, D. Nesvorný, A. Morbidelli: *Science* **294**, 1693 (2001)
10. W.F. Bottke, W.F., D. Vokrouhlický, D. P. Rubincam, M. Broz: The Effect of Yarkovsky Thermal Forces on the Dynamical Evolution of Asteroids and Meteoroids. In: *Asteroids III*, ed by W.F. Bottke JR, A. Cellino, P. Paolicchi, R.P. Binzel, T. Gehrels (Univ. of Arizona Press, Tucson 2003), pp 501–515
11. C.R. Chapman, D.R. Davis, R. Greenberg: *Meteoritics* **17**, 193 (1982)
12. C.R. Chapman, P. Paolicchi, V. Zappalà, R.P. Binzel, J.F. Bell: Asteroid families: Physical properties and evolution. In: *Asteroids II*, ed by R.P. Binzel, T. Gehrels, M.S. Matthews, (Univ. of Arizona Press, Tucson 1989), pp 386–415
13. J.S. Curtis: An accelerated reservoir light-gas gun. NASA TND-1144 (1962)
14. D.R. Davis, D.R., C.R. Chapman, S.J. Weidenschilling, R. Greenberg: *Icarus* **62**, 30 (1985)
15. D.R. Davis, E.V. Ryan: *Icarus* **83**, 156 (1990)

16. D.R. Davis, D.D. Durda, F. Marzari, A. Campo Bagatin, R. Gil-Hutton: Collisional Evolutions of Small Body Populations. In: *Asteroids III*, ed by W.F. Bottke JR, A. Cellino, P. Paolicchi, R.P. Binzel (Univ. of Arizona Press, Tucson 2003), pp 545–558
17. D.D. Durda, W.F. Bottke, B.L. Enke, W.J. Merline, E. Asphaug, D.C. Richardson, Z.M. Leinhardt: *Icarus* **170**, 243 (2004)
18. I. GIBLIN, G. Martelli, P.N. Smith, M. Di Martino: *Planet. Space Sci.* **42**, 1027 (1994)
19. D.E. Grady, M.E. Kipp: *Int. J. Rock Mech. Min. Sci. Geomech. Abstr.* **17**, 147 (1980)
20. W.K. Hartmann: *Icarus* **10**, 201 (1978)
21. K.A. Holsapple: *Plan. Space. Sci.* **42**, 1067 (1994)
22. K.R. Housen, K.A. Holsapple: *Icarus* **84**, 226 (1990)
23. K.R. Housen, R.M. Schmidt, K.A. Holsapple: *Icarus* **94**, 180 (1991)
24. Z. Knežević, A. Lemaître, A. Milani: The Determination of Asteroid Proper Elements. In: *Asteroids III*, ed by W.F. Bottke JR, A. Cellino, P. Paolicchi, R.P. Binzel (Univ. of Arizona Press, Tucson 2003), pp 603–612
25. N. Kawashima, A. Yamori, M. Yanagisawa, H. Kubo, M. Kohno, S. Teii: *IEEE Trans. Magnetics* **29**, 431 (1993)
26. B.R. Lawn, T.R. Wilshaw: *Fracture of Brittle Solids* (Cambridge University Press, New York 1975)
27. G. Martelli, P. Rothwell, P.N. Smith, I. GIBLIN, J. Martinson, E. Ducrocq, M. Wettstein, M. Di Martino, P. Farinella: Jets of fragments from catastrophic break-up and their astrophysical implications. In: *Origin and Evolution of Interplanetary Dust*, ed by H. Hasegawa, A.C. Levasseur-Regourd (Kluwer, Dordrecht 1991), pp 383–386
28. G. Martelli, E.V. Ryan, A.M. Nakamura, I. GIBLIN: *Plan. Space. Sci.* **42**, 1013 (1994)
29. F. Marzari, D. Davis, V. Vanzani: *Icarus* **113**, 168 (1995)
30. T. Matsui, T. Waza, K. Kani, S. Suzuki: *J. Geophys. Res.* **87**, 10968 (1982)
31. H.J. Melosh: *Impact cratering: a geologic process* (University Press, New York 1989)
32. H.J. Melosh, Ryan, E.V., E. Asphaug: *J. Geophys. Res.* **97**, 14735 (1992)
33. H.J. Melosh, E. V. Ryan: *Icarus* **129**, 562 (1997)
34. P. Michel, W. Benz, P. Tanga, D. C. Richardson: *Science* **294**, 1696 (2001)
35. P. Michel, W. Benz, P. Tanga, D. C. Richardson: *Icarus* **160**, 10 (2002)
36. P. Michel, W. Benz, D.C. Richardson: *Nature* **421**, 608 (2003)
37. P. Michel, W. Benz, D.C. Richardson: *Icarus* **168**, 420 (2004)
38. H. Mizutani, M. Kumazawa, M. Kato, T. Masuda, S. Kawakami, Y. Takagi, K. Kani: *Proc. ISAS Lunar Planet. Symp.* **14**, 267 (1981)
39. H. Nakamura, K. Ito, N. Fujii, H. Takeuchi, M. Miyamoto, and Y. Kobayashi: *Proc. ISAS Lunar Planet. Symp* **16**, 55 (1983)
40. A. Nakamura, A. Fujiwara: *Icarus* **92**, 132 (1991)
41. D. Nesvorný, W.F. Bottke, L. Dones, and H. F. Levison: *Nature* **417**, 720 (2002)
42. D. Nesvorný, S. Ferraz-Mello, M. Holman, A. Morbidelli: Regular and Chaotic Dynamics in the Mean-Motion Resonances: Implications for the Structure and Evolution of the Asteroid Belt. In: *Asteroids III*, ed by W.F. Bottke JR, A. Cellino, P. Paolicchi, R.P. Binzel (Univ. of Arizona Press, Tucson 2003), pp 379–394

43. P. Pravec, A.W. Harris: *Icarus* **148**, 12 (2000)
44. D.C. Richardson, D. C: *Mon. Not. R. Astron. Soc.* **269**, 493 (1994)
45. D.C. Richardson, T. Quinn, J. Stadel, and G. Lake: *Icarus* **143**, 45 (2000)
46. D.C. Richardson, Z.M. Leinhardt, W. F. Bottke, H. J. Melosh, E. Asphaug. Gravitational aggregates: Evidence and evolution. In: *Asteroids III*, ed by W.F. Bottke JR, A. Cellino, P. Paolicchi, R.P. Binzel (Univ. of Arizona Press, Tucson 2003), pp 501–515
47. E.V. Ryan, H.J. Melosh: *Icarus* **133**, 1 (1998)
48. A.E. Seigel: The theory of high speed guns. *AGARDograph* **91** (1965)
49. P. Tanga, A. Cellino, P. Michel, V. Zappalà, P. Paolicchi, and A. dell’Oro: *Icarus* **141**, 65 (1999)
50. S.L. Thompson, H.S. Lauson: Improvement in the Chart D radiation hydrodynamic code III: revised analytic equation of state. *Sandia National Laboratory Report* SC-RR-71 0714 (1972)
51. J.H. Tillotson: Metallic equations of state for hypervelocity impact. *General Atomic Report* GA-3216 (1962)
52. J.B. Walsh: *J. Geophys. Res.* **70**, 381 (1965)
53. W.R. Ward, R.M. Canup: *Nature* **403**, 741 (2000)
54. W.A. Weibull: *Ingvetensk. Akad. Handl.* **151**, 5 (1939)
55. D.K. Yeomans et al: *Science* **278**, 2106 (1997)
56. V. Zappalà, A. Cellino, A. Dell’Oro, P. Paolicchi: Physical and Dynamical Properties of Asteroid Families. In: *Asteroids III*, ed by W.F. Bottke JR, A. Cellino, P. Paolicchi, R.P. Binzel (Univ. of Arizona Press, Tucson 2003), pp 619–631

Geometric Conditions for Quasi-Collisions in Öpik's Theory

Giovanni B. Valsecchi

INAF-IASF, via Fosso del Cavaliere 100, 00133 Roma (Italy)
giovanni@rm.iasf.cnr.it

Abstract. The geometric setup of the recently proposed extension of Öpik's theory of planetary close encounters is modified in order to allow the treatment of encounters taking place far from the nodes of the small body orbit.

1 Introduction

Öpik's theory of close encounters between small bodies and planets has been successfully extended to allow the treatment of various problems related to planetary close encounters like, for instance, the “resonant returns” [4], in which a small body undergoing an encounter with a planet is perturbed by the latter into an orbit making possible a second, very close encounter, possibly a collision, at a later epoch. To extend the theory, Valsecchi et al. [7] started from the formulation by Greenberg et al. [3] and by Carusi et al. [1] of the original theory due to Öpik [5].

Öpik's original expressions relate the three components of the planetocentric unperturbed velocity vector of the small body $\mathbf{U}$ to the orbital elements a , e , i . In [3] and [1] the b -plane coordinates ξ and ζ were introduced, and were put in relation with the final (post-encounter) values of a , e , i ; here, the b -plane, often called “target plane”, is the plane containing the planet and perpendicular to the incoming asymptote of the planetocentric hyperbola that approximates the motion of the small body in the immediate vicinity of the planet, and the directions of ξ and ζ will be specified later.

Strictly speaking, Öpik's formulae are valid only at collision; [7] introduced corrections, to first order in the miss distance, that extend them to quasi-collisions (i.e., close encounters). Another contribution by [7] consists of the establishment of explicit expressions to compute ξ and ζ from the orbital elements of the small body. This is done by computing all the relevant quantities at a reference time corresponding to the passage of the small body through the node closer to where the encounter with the planet takes place; [7] show that, in this way, the quantities characterizing a planetary encounter can

be linked to those of a previous one and, in particular, that it is possible to compute the location, size and shape of the small regions in the b -plane of an encounter, the so-called “keyholes” [2], such that, if the small body passes through one of them, an impact with the planet takes place at the following encounter.

An earlier version of the theory was presented by Valsecchi [6]; the present paper, however, is based on the version given in [7].

If the inclination of the small body orbit is sufficiently small, it is not always true that very close encounters, and possibly collisions, take place close to the location of one of the nodes. In Sect. 2 the geometric setup introduced by [7] is discussed in this context, showing its shortcomings for very low inclination orbits. Then, a generalized geometric setup is introduced in Sect. 3, and its implications are discussed in Sect. 4.

2 The Geometry of Planetary Close Encounters

In order to discuss the geometry of planetary encounters, let us consider a model as compatible as possible with the assumptions made in [7]; we thus consider the Earth moving on a circular orbit of radius 1 AU, with mean motion equal to that of a massless particle.

We take, as first example, the orbit of asteroid 1999 AN₁₀, with ω slightly modified to allow very close encounters with the Earth in the vicinity of the ascending node of the asteroid orbit; this is necessary because the real 1999 AN₁₀ passes very close, at the ascending node, to the *real* orbit of the Earth, but not to its circular approximation. Thus, the orbit of “our” 1999 AN₁₀ has the following elements: $a = 1.45866$ AU, $e = 0.562104$, $i = 39^\circ 935$, $\omega = 269^\circ 717$, $\Omega = 314^\circ 497$.

The values of Öpik’s variables U , θ and ϕ (their definitions will be given in a moment) can be computed from Öpik’s formulae:

$$\begin{aligned}
 U &= \sqrt{3 - \frac{1}{a} - 2\sqrt{a(1-e^2)} \cos i} \\
 \theta &= \arccos \frac{\sqrt{a(1-e^2)} \cos i - 1}{\sqrt{3 - \frac{1}{a} - 2\sqrt{a(1-e^2)} \cos i}} \\
 \phi &= \arcsin \pm \frac{\sqrt{2 - \frac{1}{a} - a(1-e^2)}}{\sqrt{2 - \frac{1}{a} - a(1-e^2) \cos^2 i}} \\
 &= \arccos \pm \frac{\sqrt{a(1-e^2)} \sin i}{\sqrt{2 - \frac{1}{a} - a(1-e^2) \cos^2 i}},
 \end{aligned}$$

where the upper sign for $\sin \phi$ applies to encounters in the post-perihelion branch of the orbit, and the upper sign for $\cos \phi$ to encounters at the ascending

node; U is in units of the orbital velocity of the Earth. In the present case (post-perihelion, ascending node encounter) we have $U = 0.885$, $\theta = 105^\circ 3$, $\phi = 41^\circ 3$.

We then start the asteroid and the Earth along their unperturbed orbits, looking for a close approach near the ascending node of the asteroid orbit; once we find a reasonably close one, for a certain pair of mean anomalies of the two bodies, we record it, and then change only the initial mean anomaly of the Earth, by $0^\circ 2$, looking for another one, and so on.

In this way, the data in Table 1 have been obtained; in each row we have, for the time of closest approach along the unperturbed orbit:

- the longitude of the Earth $\lambda_\oplus$, in degrees;
- the impact parameter b (this is nothing else than the closest approach distance), in Earth radii;
- the coordinates on the b -plane ξ and ζ , in Earth radii;
- the values of U , θ and ϕ computed from the actual geocentric velocity vector of the asteroid (and not, as done above, from the orbital elements through Öpik's formulae);
- the geocentric coordinates X , Y , Z of the asteroid, in Earth radii.

Table 1. Various quantities related to the geometry of encounter with the Earth of 1999 AN₁₀; for details, see the text.

$\lambda_\oplus$	b	ξ	ζ	U	θ	ϕ	X	Y	Z
314° 450	18.9	9.8	-16.1	0.884	105° 4	41° 3	10.2	15.5	-3.3
314° 464	12.8	9.8	-8.2	0.884	105° 4	41° 3	8.8	7.9	-4.9
314° 478	9.8	9.8	-0.3	0.885	105° 4	41° 3	7.4	0.3	-6.4
314° 492	12.4	9.8	7.6	0.885	105° 4	41° 3	6.1	-7.3	-8.0
314° 506	18.3	9.8	15.5	0.885	105° 4	41° 3	4.7	-14.9	-9.6

The geocentric frame used is such that the Sun is on the negative X -axis, and the positive Y -axis coincides with the instantaneous direction of motion of the Earth; then, U is the magnitude of the geocentric velocity vector $\mathbf{U}$ of the asteroid, θ is the angle between $\mathbf{U}$ and the positive Y -axis, while ϕ is the angle between the planes $\mathbf{U}$ - Y and Y - Z . Thus, the three components of $\mathbf{U}$ are given by:

$$\begin{pmatrix} U_x \\ U_y \\ U_z \end{pmatrix} = \begin{pmatrix} U \sin \theta \sin \phi \\ U \cos \theta \\ U \sin \theta \cos \phi \end{pmatrix}.$$

If both θ and ϕ are equal to 0, the ξ - ζ plane, i.e. the b -plane, coincides with the X - Z plane; otherwise, to pass from the X - Y - Z frame to the b -plane frame one has to make a rotation by $-\phi$ (i.e., clockwise) about Y , followed by a rotation by $-\theta$ about the new X -axis, that has become the ξ -axis.

As noted in [7], this way of choosing the axes on the b -plane has the nice property that differences along ζ correspond to differences in the timing of the encounter, while differences along ξ correspond to different values of the local minimum of the distance between the orbits of the asteroid and that of the planet (usually indicated with the acronym MOID, for (Minimum Orbital Intersection Distance), a quantity that in general varies slowly. Thus, the encounters of a number of small bodies moving along the same orbit, and spaced in time, will be represented on the b -plane by points having the same value of ξ and different values of ζ .

This is exactly what we see for 1999 AN₁₀ in Table 1, where the values of ξ correspond to what one obtains using the expression, given in [7], that relates the local MOID to the orbital elements a , e , ω and to ϕ :

$$\xi = \left[\frac{a(1-e^2)}{1+e\cos\omega} - 1 \right] \cos\phi, \quad (1)$$

and that gives $\xi = 9.8$ Earth radii; moreover, we note that positive values of Y correspond to negative values of ζ , and that the values of U , θ and ϕ computed with Öpik's formulae approximate very well those in the Table, that are computed using the actual unperturbed geocentric velocity vector. Moreover, the values of $\lambda_\oplus$ indicate that the encounter of 1999 AN₁₀ takes place very close to the ascending node of the asteroid orbit.

The data in Table 2 refer to (4179) Toutatis, whose orbital elements are $a = 2.51117$ AU, $e = 0.633636$, $i = 0^\circ 47'$, $\omega = 274^\circ 812'$, $\Omega = 128^\circ 205'$, and whose Öpik variables are $U = 0.387$, $\theta = 54^\circ 3'$, $\phi = 88^\circ 2'$.

Table 2. Same as Table 1 for (4179) Toutatis.

$\lambda_\oplus$	b	ξ	ζ	U	θ	ϕ	X	Y	Z
80°401	142.9	142.2	-13.8	0.387	54°2'	89°1'	-5.9	11.2	-142.3
80°452	142.4	142.2	-7.2	0.387	54°2'	89°1'	-2.0	5.8	-142.3
80°502	142.2	142.2	-0.5	0.387	54°3'	89°1'	1.9	0.4	-142.2
80°552	142.4	142.2	6.1	0.388	54°4'	89°1'	5.8	-5.0	-142.2
80°602	142.8	142.2	12.8	0.388	54°5'	89°1'	9.6	-10.4	-142.1

The values of $\lambda_\oplus$ show that in this case, at variance with that of 1999 AN₁₀, the encounter takes place far away from either node; of course, this is made possible by the very low inclination, of less than half of a degree. As a consequence, the actual value of ξ , 142.2 Earth radii, differs considerably from the one computed using (1), that amounts to 320.0 Earth radii.

Such a large difference shows the inadequacy of the theory to treat cases like this one. Moreover, the considerable difference (about 48°) between Ω and $\lambda_\oplus$, that translates into a considerable difference between the time of node crossing and the time of closest approach, puts into serious question the rectilinear motion approximation used in Öpik's theory, as the computation

of the timing of sequences of encounters, in particular for resonant returns, could consequently be unreliable.

The good agreement between the actual values of U , θ and ϕ and those computed using Öpik's formulae should not be overestimated; the agreement is due to the fact that Öpik's formulae just assume that the heliocentric distance of the small body is very close to 1, an assumption well verified in cases like this, as evidenced by the values of X in Table 2.

As already said, the reason for the shortcomings of the theory is the large time difference between the passage at the node by the small body and its encounter with the planet. To solve this problem, a version of the theory in which the reference time is not necessarily the time of node crossing is developed in the next section.

3 A Generalized Setup for Öpik's Theory

In the X - Y - Z frame, if t_0 is a generic reference time at which the distance between the small body and the planet is small enough that terms beyond the first order in the distance between the small body and the planet can be ignored, we can write:

$$\begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix} = \begin{pmatrix} U_x(t-t_0) + X(t_0) \\ U_y(t-t_0) + Y(t_0) \\ U_z(t-t_0) + Z(t_0) \end{pmatrix},$$

where $X(t_0) = X_0$, $Y(t_0) = Y_0$ and $Z(t_0) = Z_0$ are the planetocentric coordinates of the small body at time t_0 ; thus, the rectilinear motion of the small body close to the planet is described by:

$$\begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix} = \begin{pmatrix} U_x(t-t_0) + X_0 \\ U_y(t-t_0) + Y_0 \\ U_z(t-t_0) + Z_0 \end{pmatrix} = \begin{pmatrix} U(t-t_0) \sin \theta \sin \phi + X_0 \\ U(t-t_0) \cos \theta + Y_0 \\ U(t-t_0) \sin \theta \cos \phi + Z_0 \end{pmatrix}.$$

3.1 From Heliocentric Elements of the Small Body to Cartesian Geocentric Position and Velocity and Back

Expressions for the computation of a , e , i from U , θ , ϕ are given in many papers (see, e.g., [1, 6, 7]); those for the computation of U , θ , ϕ from a , e , i , provided we know the sign of U_x and U_z , were given in the previous section.

For the remaining quantities, let us call r the heliocentric distance, and λ , β the longitude and latitude of the small body at time t_0 ; these quantities are given by:

$$\begin{aligned} r &= \frac{a(1-e^2)}{1+e \cos f} \\ \lambda &= \Omega + \arctan[\cos i \tan(\omega + f)] \\ \beta &= \arcsin(\sin i \sin(\omega + f)). \end{aligned}$$

At time t_0 , with the planet at longitude $\lambda_\oplus$, we have:

$$\begin{aligned} X_0 &= r \cos(\lambda - \lambda_\oplus) \cos \beta - 1 \\ Y_0 &= r \sin(\lambda - \lambda_\oplus) \cos \beta \\ Z_0 &= r \sin \beta . \end{aligned}$$

Since at time t_0 the geocentric distance is small, then $|r - 1| \ll 1$, $\lambda - \lambda_\oplus \ll 1$ and $\beta \ll 1$, so that we can keep only terms of the first order in these quantities and write:

$$\begin{aligned} X_0 &= r - 1 \\ &= \frac{a(1 - e^2)}{1 + e \cos f} - 1 \\ Y_0 &= \lambda - \lambda_\oplus \\ &= \Omega + \arctan[\cos i \tan(\omega + f)] - \lambda_\oplus \\ Z_0 &= \sin i \sin(\omega + f) . \end{aligned}$$

We still have some freedom in the definition of t_0 ; for example, we can define it as the time when the small body is at one of its nodes, as done in [7]. In this case, either $\omega + f = 0$, at the ascending node, or $\omega + f = 180^\circ$, at the descending node. We have then:

$$\begin{aligned} X_0 &= \frac{a(1 - e^2)}{1 \pm e \cos \omega} - 1 \\ Y_0 &= \Omega + 90^\circ \mp 90^\circ - \lambda_\oplus \\ Z_0 &= 0 , \end{aligned}$$

with the upper sign applying at the ascending node, and the lower sign at the descending one. These relationships are easily inverted, to get ω , Ω and f from X_0 and Y_0 :

$$\begin{aligned} \Omega &= \lambda_\oplus + \arctan \frac{Y_0}{1 + X_0} + 90^\circ \mp 90^\circ \\ \cos \omega &= \pm \frac{a(1 - e^2) - (1 + X_0)}{e(1 + X_0)} \\ f &= 90^\circ \mp 90^\circ - \omega . \end{aligned}$$

Other choices are possible; for example, defining t_0 such that $\lambda - \lambda_\oplus = 0$, we have:

$$\begin{aligned} X_0 &= \frac{a(1 - e^2)}{1 + e \cos f} - 1 \\ Y_0 &= 0 \\ Z_0 &= \sin i \sin(\omega + f) . \end{aligned}$$

If instead we choose t_0 such that $r = 1$, we have:

$$\begin{aligned} X_0 &= 0 \\ Y_0 &= \Omega + \arctan[\cos i \tan(\omega + f)] - \lambda_{\oplus} \\ Z_0 &= \sin i \sin(\omega + f) . \end{aligned}$$

In the latter two cases, the passage from X_0, Y_0, Z_0 to ω, Ω, f is less straightforward, but can be done with the usual procedures giving orbital elements from cartesian coordinates and velocities.

3.2 The Local MOID

We now proceed to establish an expression for the local MOID. In

$$\begin{pmatrix} X(t) \\ Y(t) \\ Z(t) \end{pmatrix} = \begin{pmatrix} U(t - t_0) \sin \theta \sin \phi + X_0 \\ U(t - t_0) \cos \theta + Y_0 \\ U(t - t_0) \sin \theta \cos \phi + Z_0 \end{pmatrix}$$

we eliminate $t - t_0$, using:

$$t - t_0 = \frac{Y - Y_0}{U \cos \theta} ,$$

and obtain:

$$\begin{aligned} X &= (Y - Y_0) \tan \theta \sin \phi + X_0 \\ Z &= (Y - Y_0) \tan \theta \cos \phi + Z_0 . \end{aligned}$$

Setting $w = Y - Y_0$, the square of the distance from the Y -axis is:

$$\begin{aligned} D_y^2 &= X^2 + Z^2 \\ &= w^2 \tan^2 \theta + 2w(X_0 \sin \phi + Z_0 \cos \phi) \tan \theta + X_0^2 + Z_0^2 \end{aligned}$$

and its derivative with respect to w is:

$$\frac{d(D_y^2)}{dw} = 2w \tan^2 \theta + 2(X_0 \sin \phi + Z_0 \cos \phi) \tan \theta ;$$

this derivative is zero at:

$$w_{MOID} = -(X_0 \sin \phi + Z_0 \cos \phi) \cot \theta .$$

The minimum value of D_y^2 is then:

$$\min D_y^2 = (X_0 \cos \phi - Z_0 \sin \phi)^2 .$$

The local MOID as function of X_0, Z_0 and ϕ is:

$$\min D_y = |X_0 \cos \phi - Z_0 \sin \phi| ;$$

following [7], we define the signed local MOID as $X_0 \cos \phi - Z_0 \sin \phi$.

We can now compute all the relevant quantities when the small body encounters the planet at a minimum distance equal to the local MOID. For Y_{MOID} we have:

$$Y_{MOID} = Y_0 - (X_0 \sin \phi + Z_0 \cos \phi) \cot \theta ,$$

and for t_{MOID} , X_{MOID} and Z_{MOID} we have:

$$\begin{aligned} t_{MOID} &= t_0 - \frac{X_0 \sin \phi + Z_0 \cos \phi}{U \sin \theta} \\ X_{MOID} &= X_0 - (X_0 \sin \phi + Z_0 \cos \phi) \sin \phi \\ Z_{MOID} &= Z_0 - (X_0 \sin \phi + Z_0 \cos \phi) \cos \phi . \end{aligned}$$

When the small body is at the local MOID, its Y -coordinate must be 0, as the segment corresponding to the minimum distance between two straight lines must be on the straight line that is the intersection of the planes perpendicular to each of the given straight lines, and in particular it must be in the X - Z -plane. To check this, we compute explicitly the local MOID:

$$\begin{aligned} \min D_y^2 &= X_{MOID}^2 + Y_{MOID}^2 + Z_{MOID}^2 \\ &= (X_0 \cos \phi - Z_0 \sin \phi)^2 + [Y_0 - (X_0 \sin \phi + Z_0 \cos \phi) \cot \theta]^2 . \end{aligned}$$

Since we must have $\min D_y^2 = (X_0 \cos \phi - Z_0 \sin \phi)^2$, it follows that the condition is

$$Y_0 = (X_0 \sin \phi + Z_0 \cos \phi) \cot \theta ,$$

that implies

$$Y_{MOID} = Y_0 - (X_0 \sin \phi + Z_0 \cos \phi) \cot \theta = 0 .$$

Thus,

$$\begin{pmatrix} X_{MOID} \\ Y_{MOID} \\ Z_{MOID} \end{pmatrix} = \begin{pmatrix} X_0 - (X_0 \sin \phi + Z_0 \cos \phi) \sin \phi \\ 0 \\ Z_0 - (X_0 \sin \phi + Z_0 \cos \phi) \cos \phi \end{pmatrix} ,$$

and we can compute the coordinates on the b -plane using the formulae given in [7]:

$$\begin{aligned} \xi &= X_{MOID} \cos \phi - Z_{MOID} \sin \phi \\ &= X_0 \cos \phi - Z_0 \sin \phi \\ \eta &= (X_{MOID} \sin \phi + Z_{MOID} \cos \phi) \sin \theta + Y_{MOID} \cos \theta \\ &= 0 \\ \zeta &= (X_{MOID} \sin \phi + Z_{MOID} \cos \phi) \cos \theta - Y_{MOID} \sin \theta \\ &= 0 . \end{aligned}$$

3.3 The Coordinates on the b -Plane

We now proceed in a similar way to determine the coordinates on the b -plane in the general case in which, at $t = t_0$, the small body is at a generic point (X_0, Y_0, Z_0) not necessarily leading to an encounter at the MOID; we then have:

$$\begin{aligned} X &= U \sin \theta \sin \phi(t - t_0) + X_0 \\ Y &= U \cos \theta(t - t_0) + Y_0 \\ Z &= U \sin \theta \cos \phi(t - t_0) + Z_0 \end{aligned}$$

and we want to minimize the distance from the planet:

$$\begin{aligned} D^2 &= X^2 + Y^2 + Z^2 \\ &= U^2 t^2 + 2U[(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + Y_0 \cos \theta - Ut_0]t \\ &\quad - 2U[(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + Y_0 \cos \theta - Ut_0]t_0 + X_0^2 + Y_0^2 + Z_0^2. \end{aligned}$$

We take the derivative with respect to t :

$$\frac{d(D^2)}{dt} = 2U^2 t + 2U[(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + Y_0 \cos \theta - Ut_0]$$

and find the value $t = t_b$ for which it is zero:

$$t_b = t_0 - \frac{(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + Y_0 \cos \theta}{U}. \quad (2)$$

Thus, one has the minimum approach distance when the small body is in:

$$\begin{aligned} X_b &= U \sin \theta \sin \phi(t_b - t_0) + X_0 \\ &= X_0 - [(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + Y_0 \cos \theta] \sin \theta \sin \phi \\ Y_b &= U \cos \theta(t_b - t_0) + Y_0 \\ &= Y_0 - [(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + Y_0 \cos \theta] \cos \theta \\ Z_b &= U \sin \theta \cos \phi(t_b - t_0) + Z_0 \\ &= Z_0 - [(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + Y_0 \cos \theta] \sin \theta \cos \phi. \end{aligned}$$

The coordinates on the b -plane are:

$$\begin{aligned} \xi &= X_b \cos \phi - Z_b \sin \phi \\ &= X_0 \cos \phi - Z_0 \sin \phi \\ \eta &= (X_b \sin \phi + Z_b \cos \phi) \sin \theta + Y_b \cos \theta \\ &= 0 \\ \zeta &= (X_b \sin \phi + Z_b \cos \phi) \cos \theta - Y_b \sin \theta \\ &= (X_0 \sin \phi + Z_0 \cos \phi) \cos \theta - Y_0 \sin \theta. \end{aligned}$$

In general, given θ and ϕ , the coordinates ξ and ζ on the b -plane depend on X_0 , Y_0 and Z_0 ; if we want to compute X_0 , Y_0 and Z_0 from ξ and ζ , we need to know $t_b - t_0$. Inverting (2) we can compute Y_0 as function of X_0 and Z_0 :

$$Y_0 = -\frac{(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + U(t_b - t_0)}{\cos \theta} ;$$

then, from the expression giving ξ as function of X_0 and Z_0 we compute Z_0 as function of X_0 and ξ :

$$Z_0 = \frac{X_0 \cos \phi - \xi}{\sin \phi} .$$

We then substitute Y_0 and Z_0 in the expression giving ζ as function of X_0 , Y_0 and Z_0 , and solve for X_0 :

$$\begin{aligned} \zeta &= (X_0 \sin \phi + Z_0 \cos \phi) \cos \theta - Y_0 \sin \theta \\ &= X_0[(\cos \theta + \sin \theta \tan \theta)(\sin \phi + \cos \phi \cot \phi)] \\ &\quad - \xi(\cos \theta + \sin \theta \tan \theta) \cot \phi + U(t_b - t_0) \tan \theta \\ X_0 &= [\zeta \cos \theta - U(t_b - t_0) \sin \theta] \sin \phi + \xi \cos \phi ; \end{aligned}$$

finally, we get explicit expressions for Z_0 and Y_0 :

$$\begin{aligned} Z_0 &= [\zeta \cos \theta - U(t_b - t_0) \sin \theta] \cos \phi - \xi \sin \phi \\ Y_0 &= -(\zeta \sin \theta + U(t_b - t_0) \cos \theta) . \end{aligned}$$

3.4 The Encounter

At time

$$t_b = t_0 - \frac{(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta + Y_0 \cos \theta}{U} ,$$

corresponding to the b -plane crossing, we rotate the velocity vector, that is parallel to the incoming asymptote of the geocentric hyperbola, so as to make it parallel to the other asymptote; moreover, the position of the small body becomes the one corresponding to the minimum unperturbed distance on the new orbit; the coordinates in the ξ - η - ζ reference frame pass from

$$\begin{pmatrix} \xi \\ \eta \\ \zeta \end{pmatrix} = \begin{pmatrix} X_0 \cos \phi - Z_0 \sin \phi \\ 0 \\ (X_0 \sin \phi + Z_0 \cos \phi) \cos \theta - Y_0 \sin \theta \end{pmatrix}$$

to

$$\begin{pmatrix} \xi_r \\ \eta_r \\ \zeta_r \end{pmatrix} = \begin{pmatrix} \xi \cos \gamma \\ b \sin \gamma \\ \zeta \cos \gamma \end{pmatrix}$$

Following [7], we define

$$c = \frac{m}{U^2},$$

and use the expressions for $\sin \gamma$ and $\cos \gamma$

$$\begin{aligned}\cos \gamma &= \frac{b^2 - c^2}{b^2 + c^2} \\ \sin \gamma &= \frac{2bc}{b^2 + c^2}\end{aligned}$$

to rewrite the previous expressions for the components of the rotated vector $\mathbf{b}$, that we call $\mathbf{b}'$, in the ξ - η - ζ reference frame

$$\begin{pmatrix} \xi_r \\ \eta_r \\ \zeta_r \end{pmatrix} = \begin{pmatrix} \frac{\xi(b^2 - c^2)}{b^2 + c^2} \\ \frac{2b^2 c}{b^2 + c^2} \\ \frac{\zeta(b^2 - c^2)}{b^2 + c^2} \end{pmatrix}.$$

We denote by X'_b , Y'_b , Z'_b the components of $\mathbf{b}'$ in the X - Y - Z frame; their explicit expressions are the following:

$$\begin{aligned}X'_b &= (\eta_r \sin \theta + \zeta_r \cos \theta) \sin \phi + \xi_r \cos \phi \\ &= \frac{2b^2 c \sin \theta \sin \phi + (b^2 - c^2)(\zeta \cos \theta \sin \phi + \xi \cos \phi)}{b^2 + c^2} \\ Y'_b &= \eta_r \cos \theta - \zeta_r \sin \theta \\ &= \frac{2b^2 c \cos \theta - (b^2 - c^2)\zeta \sin \theta}{b^2 + c^2} \\ Z'_b &= (\eta_r \sin \theta + \zeta_r \cos \theta) \cos \phi - \xi_r \sin \phi \\ &= \frac{2b^2 c \sin \theta \cos \phi + (b^2 - c^2)(\zeta \cos \theta \cos \phi - \xi \sin \phi)}{b^2 + c^2};\end{aligned}$$

the components of the rotated velocity vector $\mathbf{U}'$ are given by:

$$\begin{aligned}U'_x &= U \sin \theta' \sin \phi' \\ &= U \frac{[(b^2 - c^2) \sin \theta - 2c\zeta \cos \theta] \sin \phi - 2c\xi \cos \phi}{b^2 + c^2} \\ U'_y &= U \cos \theta' \\ &= U \frac{(b^2 - c^2) \cos \theta + 2c\zeta \sin \theta}{b^2 + c^2} \\ U'_z &= U \sin \theta' \cos \phi' \\ &= U \frac{[(b^2 - c^2) \sin \theta - 2c\zeta \cos \theta] \cos \phi + 2c\xi \sin \phi}{b^2 + c^2}.\end{aligned}$$

3.5 Post-Encounter Coordinates in the Post-Encounter $\mathbf{b}$ -Plane and the New Local MOID

We can apply the appropriate rotations by θ' and ϕ' to the components of $\mathbf{b}'$ in the X - Y - Z frame to obtain the coordinates in the post-encounter $\mathbf{b}$ -plane

(this plane is defined in the same way as the ξ - ζ plane, but is orthogonal to $\mathbf{U}'$):

$$\begin{aligned}
 \xi' &= X'_b \cos \phi' - Z'_b \sin \phi' \\
 &= \frac{(b^2 + c^2)\xi \sin \theta}{\sqrt{[(b^2 - c^2) \sin \theta - 2c\zeta \cos \theta]^2 + 4c^2\xi^2}} \\
 \eta' &= (X'_b \sin \phi' + Z'_b \cos \phi') \sin \theta' + Y'_b \cos \theta' \\
 &= 0 \\
 \zeta' &= (X'_b \sin \phi' + Z'_b \cos \phi') \cos \theta' - Y'_b \sin \theta' \\
 &= \frac{(b^2 - c^2)\zeta \sin \theta - 2b^2c \cos \theta}{\sqrt{[(b^2 - c^2) \sin \theta - 2c\zeta \cos \theta]^2 + 4c^2\xi^2}}.
 \end{aligned}$$

Note that ξ' is the new local MOID.

3.6 Post-Encounter Propagation

The coordinates at a generic time t along the post-encounter trajectory of the small body can be obtained from

$$\begin{aligned}
 X'(t) &= U'_x(t - t_b) + X'_b \\
 &= \frac{U(t - t_b)\{[(b^2 - c^2) \sin \theta - 2c\zeta \cos \theta] \sin \phi - 2c\xi \cos \phi\}}{b^2 + c^2} \\
 &\quad + \frac{2b^2c \sin \theta \sin \phi + (b^2 - c^2)(\zeta \cos \theta \sin \phi + \xi \cos \phi)}{b^2 + c^2} \\
 Y'(t) &= U'_y(t - t_b) + Y'_b \\
 &= \frac{U(t - t_b)[(b^2 - c^2) \cos \theta + 2c\zeta \sin \theta]}{b^2 + c^2} + \frac{2b^2c \cos \theta - (b^2 - c^2)\zeta \sin \theta}{b^2 + c^2} \\
 Z'(t) &= U'_z(t - t_b) + Z'_b \\
 &= \frac{U(t - t_b)\{[(b^2 - c^2) \sin \theta - 2c\zeta \cos \theta] \cos \phi + 2c\xi \sin \phi\}}{b^2 + c^2} \\
 &\quad + \frac{2b^2c \sin \theta \cos \phi + (b^2 - c^2)(\zeta \cos \theta \cos \phi - \xi \sin \phi)}{b^2 + c^2}.
 \end{aligned}$$

These expressions allow us to compute the post-encounter reference time t'_0 corresponding to one of three possibilities ($X'_0 = X'(t'_0) = 0$, $Y'_0 = Y'(t'_0) = 0$, $Z'_0 = Z'(t'_0) = 0$). In particular, for $X'_0 = X'(t'_0) = 0$:

$$t'_0 = t_b - \frac{2b^2c \sin \theta \sin \phi + (b^2 - c^2)(\zeta \cos \theta \sin \phi + \xi \cos \phi)}{U\{[(b^2 - c^2) \sin \theta - 2c\zeta \cos \theta] \sin \phi - 2c\xi \cos \phi\}};$$

for $Y'_0 = Y'(t'_0) = 0$:

$$t'_0 = t_b - \frac{2b^2c \cos \theta - (b^2 - c^2)\zeta \sin \theta}{U[(b^2 - c^2) \cos \theta + 2c\zeta \sin \theta]};$$

finally, for $Z'_0 = Z'(t'_0) = 0$ (the case developed in [7]):

$$t'_0 = t_b - \frac{2b^2 c \sin \theta \cos \phi + (b^2 - c^2)(\zeta \cos \theta \cos \phi - \xi \sin \phi)}{U\{(b^2 - c^2) \sin \theta - 2c\zeta \cos \theta\} \cos \phi + 2c\xi \sin \phi}.$$

4 Discussion

As we have already seen, we can set one of the three values (X_0, Y_0, Z_0) equal to 0, leading to simpler formulae. Which one, among these choices, is the most convenient?

A possible criterion could be established by looking at the value of $t_b - t_0$ in each case of interest, putting then equal to 0 the coordinate that gives the smallest $t_b - t_0$. We have that, for $X_0 = 0$,

$$t_b - t_0 = -\frac{Z_0 \cos \phi \sin \theta + Y_0 \cos \theta}{U};$$

for $Y_0 = 0$,

$$t_b - t_0 = -\frac{(X_0 \sin \phi + Z_0 \cos \phi) \sin \theta}{U};$$

and, for $Z_0 = 0$,

$$t_b - t_0 = -\frac{X_0 \sin \phi \sin \theta + Y_0 \cos \theta}{U}.$$

Comparing the numerical values of these three expressions, one can choose the “version” of the theory most suitable for the case at hand.

The expressions linking b -plane coordinates to coordinates in the X - Y - Z frame are as follows: for $X_0 = 0$ we have

$$\begin{aligned}\xi &= -Z_0 \sin \phi \\ \zeta &= Z_0 \cos \theta \cos \phi - Y_0 \sin \theta \\ Z_0 &= -\frac{\xi}{\sin \phi} \\ Y_0 &= -\frac{\zeta + \xi \cos \theta \cot \phi}{\sin \theta};\end{aligned}$$

for $Y_0 = 0$

$$\begin{aligned}\xi &= X_0 \cos \phi - Z_0 \sin \phi \\ \zeta &= (X_0 \sin \phi + Z_0 \cos \phi) \cos \theta \\ X_0 &= \frac{\xi \cos \theta \cot \phi + \zeta}{\cos \theta (\sin \phi + \cos \phi \cot \phi)} \\ Z_0 &= \frac{\xi \cos \theta \cot \phi + \zeta}{\cos \theta \tan \phi (\sin \phi + \cos \phi \cot \phi)} - \frac{\xi}{\sin \phi};\end{aligned}$$

finally, for $Z_0 = 0$

$$\begin{aligned}\xi &= X_0 \cos \phi \\ \zeta &= X_0 \cos \theta \sin \phi - Y_0 \sin \theta \\ X_0 &= \frac{\xi}{\cos \phi} \\ Y_0 &= \frac{\xi \cos \theta \tan \phi - \zeta}{\sin \theta} .\end{aligned}$$

Armed with this knowledge, we can now reconsider the case of (4179) Toutatis. As discussed in Sect. 2, there is a large discrepancy between the actual value of $\xi = 142.2$ Earth radii given in Table 2 and the value than one obtains from $X_0 \cos \phi = 320.0$ Earth radii, computed for the case in which, at time t_0 , the asteroid crosses the ecliptic (i.e., $Z_0 = 0$).

We now use the formulae that apply when, at time t_0 , the heliocentric distance of the asteroid is 1 AU (i.e., $X_0 = 0$); in this case, $Z_0 = -142.3$ Earth radii, and $-Z_0 \sin \phi = 142.2$ Earth radii, in very good agreement with Table 2.

As final remark, it is worth to point out that, starting from the version of the theory in which $X_0 = 0$, it is rather straightforward to build a 2-dimensional Öpik's theory of close encounters, in which the b -plane reduces to the ζ -axis.

Acknowledgments

Useful discussions on the geometry of the problem with S.R. Chesley, G.F. Gronchi, G. Tommei, and especially with A. Milani, are gratefully acknowledged.

References

1. A. Carusi, G.B. Valsecchi and R. Greenberg: *Celest. Mech. & Dynam. Astron.* **49**, 111 (1990)
2. P.W. Chodas: *Bull. Amer. Astron. Soc.* **31**, 1117 (1999)
3. R. Greenberg, A. Carusi and G.B. Valsecchi: *Icarus* **75**, 1 (1988)
4. A. Milani, S.R. Chesley and G.B. Valsecchi: *Astron. Astrophys.* **346**, L65 (1999)
5. E.J. Öpik: *Interplanetary Encounters*, (Elsevier, New York 1976)
6. G.B. Valsecchi: Close encounters in Öpik's theory. In: *Singularities in gravitational systems. Applications to Chaotic Transport in the Solar System*, ed. by Cl. Froeschlé, D. Benest (Springer, Berlin Heidelberg New York 2002) pp. 145–178
7. G.B. Valsecchi, A. Milani, G.F. Gronchi and S.R. Chesley: *Astron. Astrophys.* **408**, 1179 (2003)

The Synchronous Rotation of the Moon

Jacques Henrard

Département de mathématique FUNDP
8, Rempart de la Vierge, B-5000 Namur, Belgique

1 Introduction

Like most of the regular satellites of the solar system, the Moon presents the same face to its planet. Reference [9] showed how this peculiar feature correspond to an equilibrium, a *Cassini's state*, of a simplified model of the rotation and how perturbations from this model lead not to the destabilization of this equilibrium but to the excitation of *librations* around it. The fact that so many satellites are found in this special state is due to internal dissipations of energy in the satellites which drive them to a state of minimum energy [4]. For the Moon itself several more or less complete theories of the librations around the mean equilibrium have been developed. Let us mention the works of Eckhardt (1965), Migus (1980) and [7]. The last one with its complements [8] seems the most elaborate.

The dynamical significance of this spin-orbit resonance has been further explored and generalized ([9]; [5]). It was found that the equilibrium described by Cassini is not the only quasi equilibrium state, but that one or three others (depending on the values of the parameters) are possible.

The aim of this paper is to provide a rather complete analysis of the Cassini's states in the case of the Moon, but of course on a simplified model. We will assume that the orbit of the Moon is a circular orbit. This model may seem to be rather crude; but as we shall see it is sufficient to describe the main features of the synchronous rotation. The consideration of a perturbed elliptic orbit, or of the perturbations due to the Sun or the planets, as was done by the authors mentioned above, would, of course, lead to a more refined theory; but the analysis of a simple model with fewer parameters brings forward, in all clarity, the important features of the problem.

2 Andoyer's Variables

We will use, as most of the modern theories of rigid bodies rotation, Andoyer's variables ([1] – see also [3]) to describe the dynamics of a rigid body around its center of mass. They consists in two linked sets of Euler's angles (see Fig. 1); a first set locates the position of the angular momentum in an inertial frame of reference; the second locates the body frame in the previous frame tied to the angular momentum.

The canonical set of Andoyer's variables consists in the three angular variables ℓ, g, h and their conjugated momenta defined by the norm G of the angular momentum and two of its projections:

$$\begin{aligned} \ell &, & L &= G \cos J, \\ g &, & G &, \\ h &, & H &= G \cos K. \end{aligned}$$

With these variables the vectors $\vec{\omega}$ (the instantaneous rotation vector) and $\vec{G}$ (the angular momentum vector with respect to the center of mass) can be easily computed. their components in the frame of the principal axis of the body are:

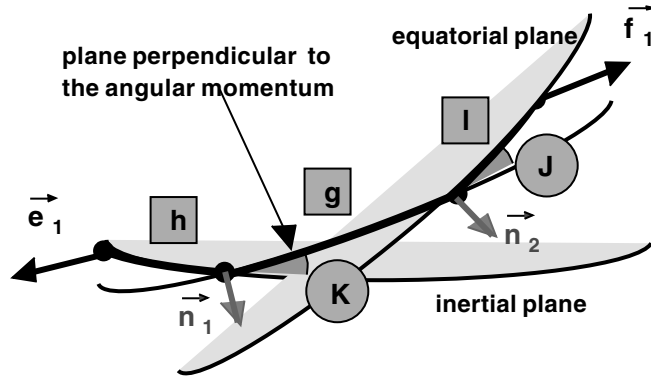


Fig. 1. The linked sets of Euler angles (ℓ, Kg) and (g, J, h) , from which are defined the Andoyer's angular variables. They locate the body frame $(\vec{f}_1, \vec{f}_2, \vec{f}_3)$ with respect to the inertial frame $(\vec{e}_1, \vec{e}_2, \vec{e}_3)$

$$\begin{aligned} \vec{\omega} &= (A^{-1}G \sin J \sin \ell, B^{-1}G \sin J \cos \ell, C^{-1}G \cos J), \\ \vec{G} &= (G \sin J \sin \ell, G \sin J \cos \ell, G \cos J), \end{aligned}$$

where $(A \leq B \leq C)$ are the principal moment of inertia of the body. The kinetic energy of the rotation is thus:

$$\begin{aligned}
T = \frac{1}{2}(\vec{\omega} | \vec{L}_C) &= \frac{1}{2}G^2 \sin^2 J \left[\frac{\sin^2 \ell}{A} + \frac{\cos^2 \ell}{B} \right] + \frac{G^2 \cos^2 J}{2C}, \\
&= \frac{1}{2}(G^2 - L^2) \left[\frac{\sin^2 \ell}{A} + \frac{\cos^2 \ell}{B} \right] + \frac{L^2}{2C},
\end{aligned} \tag{1}$$

which is the Hamiltonian of the free body rotation. Notice that the only angular variable appearing in it is the variable ℓ . Hence the problem is reduced to a one-degree-of freedom problem the phase space of which is described in [3] and in [7].

The Andoyer's variables present so-called *virtual singularities*; When $J = 0$ the angular variables ℓ and g are undefined but their sum is well defined; when $K = 0$ it is the angles g and h which are not defined although their sum is well defined. In our case it is only the first of these virtual singularities which makes problem. We shall thus use the following *modified Andoyer's variables*:

$$\begin{aligned}
p &= \ell + g + h, & P &= G, \\
r &= -h, & R &= G - H = G(1 - \cos K) = 2G \sin^2 \frac{K}{2}, \\
\xi_q &= \sqrt{2Q} \sin q, & \eta_q &= \sqrt{2Q} \cos q,
\end{aligned}$$

where $q = -\ell$ and $Q = G - L = G(1 - \cos J) = 2G \sin^2 J/2$. The free rotation Hamiltonian is now written:

$$\mathcal{H} = \frac{P^2}{2C} + \frac{4P - \xi_q^2 - \eta_q^2}{8C} \left[\frac{\alpha + \beta}{1 - \alpha - \beta} \xi_q^2 + \frac{\alpha - \beta}{1 - \alpha + \beta} \eta_q^2 \right], \tag{2}$$

where:

$$\begin{aligned}
\alpha &= (2C - A - B)/2C = J_2 MR^2/C \approx 0.4 J_2 \approx 5.15 \cdot 10^{-4} \\
\beta &= (B - A)/2C = 2C_{22} MR^2/C \approx 0.8 C_{22} \approx 1.15 \cdot 10^{-4},
\end{aligned} \tag{3}$$

3 Perturbation by Another Body

Assume that a punctual mass m_p perturbs the rotation of the body. The gravitational potential of the perturbation can be written:

$$V = -Gm_p \iiint_W \frac{\rho dW}{r'},$$

where ρ is the density inside the volume W of the rigid body under consideration and r' is the distance between the punctual mass and a volume element inside the body.

Using the usual expansion of the potential in spherical harmonics, we find

$$V = -\frac{Gm_p}{r} \left\{ 1 + \sum_{n \geq 1} \frac{1}{r^n} \sum_{m=0}^n P_n^m(\sin \phi) [C_n^m \cos m\lambda + S_n^m \sin m\lambda] \right\},$$

where λ and ϕ are the longitude and latitude of the perturbing mass in the body frame. If we limit the expansion to the second order terms and neglect the first term Gm_p/r which does not produce any effect on the rotation, we have:

$$V = -\frac{3Gm_p}{2r^3} [J_2(x^2 + y^2) + 2C_{22}(x^2 - y^2)] , \quad (4)$$

where (x, y, z) are the component (in the body frame) of the unit vector pointing to the perturbing body. We shall assume that the perturbing body is on a (possibly perturbed) Keplerian orbit around the rigid body. We may thus introduce the Keplerian mean motion $n = \sqrt{(M + m_p)/a^3}$ and the parameters

$$\begin{aligned} \delta_1 &= -\frac{3}{2} \frac{m_p}{m_p + M} \alpha \approx -3.816 \cdot 10^{-4} \\ \delta_2 &= -\frac{3}{2} \frac{m_p}{m_p + M} \beta \approx -0.852 \cdot 10^{-4} . \end{aligned} \quad (5)$$

With these notations, the potential reads:

$$V = n^2 C \left(\frac{a}{r} \right)^3 [\delta_1(x^2 + y^2) + \delta_2(x^2 - y^2)] . \quad (6)$$

From thereon, we shall assume that the perturbing body is on a slowly precessing circular orbit, with inclination i and precession frequency $\dot{\Omega}$. Taking the eccentricity into account, or for that matter taking a more general perturbed orbit into account, would not be difficult, but would introduce many supplementary parameters and we would like to keep the analysis as simple and as general as possible.

The time dependence of the Hamiltonian can be modelled by introducing two new angular variables, the longitude $\lambda = M + \omega + \Omega = nt + \lambda_0$ of the perturber and its node $\Omega = \dot{\Omega}t + \Omega_0$, and their conjugated momenta Λ and Φ . Also we wish to put forward the angular variables $\sigma = p - \lambda$ and $\rho = r + \Omega$, which will turn out to be resonances variables. We then introduce the set of canonical variables:

$$\begin{aligned} \sigma &= p - \lambda , & S &= P \\ \rho &= r + \Omega , & R &= R \\ \xi_q &= \xi_q , & \eta_q &= \eta_q \\ \lambda &= \lambda , & \Lambda' &= \Lambda + P \\ \Omega &= \Omega , & \Phi' &= \Phi - R \end{aligned} \quad (7)$$

The Hamiltonian then becomes:

$$\begin{aligned} \mathcal{H} &= n\Lambda' + \frac{S^2}{2C} - nS + \dot{\Omega}(\Phi' + R) \\ &+ \frac{4S - \xi_q^2 - \eta_q^2}{8C} \left[\frac{\alpha + \beta}{1 - \alpha - \beta} \xi_q^2 + \frac{\alpha - \beta}{1 - \alpha + \beta} \eta_q^2 \right] \\ &+ n^2 C [\delta_1(x^2 + y^2) + \delta_2(x^2 - y^2)] . \end{aligned}$$

4 Cassini's States

The rotation period of the Moon (and of other satellites like the Galilean satellites) is equal to its orbital period. Also, as Cassini pointed out, the node of the orbit and the node of the equator have, in the mean, the same period. There is thus a *resonance* between the angles λ and p on one hand, and the angles Ω and $-r$. This is why we have defined in (7) the two *resonance angular variables* σ and ρ .

As S/C (the first approximation of the angular velocity of the satellite) is almost equal to n (the angular velocity of the orbit), there is only one *fast* variable, the angular variable λ , in the list (7). A first step in the theory of the rotation of the satellite is to perform an *averaging* canonical transformation in order to eliminate this variable from the expression of the Hamiltonian. At the first order the effect of the averaging transformation is simply to remove all the terms which contain this variable. We assume that this step has been performed and obtain the *averaged* Hamiltonian:

$$\begin{aligned}\mathcal{H} = n\Lambda' + \frac{S^2}{2C} - nS + \dot{\Omega}(\Phi' + R) \\ \frac{S}{2C} \left[\frac{\alpha+\beta}{1-\alpha-\beta} \xi_q^2 + \frac{(\alpha-\beta)}{1-\alpha+\beta} \eta_q^2 \right] \\ + n^2 C [\delta_1 \mathcal{S} + \delta_2 \mathcal{D}] .\end{aligned}$$

We have neglected the terms of degree 4 in ξ_q and η_q , as the next step in the theory will anyway discard them. The functions $\mathcal{S}$ and $\mathcal{D}$ are:

$$\begin{aligned}\mathcal{S} = -\frac{1}{4}(2 - si^2) + \frac{1}{8}sk^2(2 - 3si^2) + sicskck \cos \rho + \frac{1}{4}si^2sk^2 \cos 2\rho \\ -(\xi_q^2 + \eta_q^2) \left\{ \begin{array}{l} \frac{1}{8}(2 - 3si^2)(2 - 3sk^2) \\ + \frac{3}{2}sicskck \cos \rho + \frac{3}{8}si^2sk^2 \cos 2\rho \end{array} \right. \\ -(\xi_q^2 - \eta_q^2) \left\{ \begin{array}{l} \frac{1}{32}(1 + ci)^2(1 + ck)^2 \cos 2\sigma \\ + \frac{1}{8}sisk(1 + ci)(1 + ck) \cos(2\sigma + \rho) \\ + \frac{3}{16}si^2sk^2 \cos(2\sigma + 2\rho) \\ + \frac{1}{8}sisk(1 - ci)(1 - ck) \cos(2\sigma + 3\rho) \\ + \frac{1}{32}(1 - ci)^2(1 - ck)^2 \cos(2\sigma + 4\rho) , \end{array} \right.\end{aligned}$$

and:

$$\mathcal{D} = (2 - \xi_q^2 - \eta_q^2) \left\{ \begin{array}{l} \frac{1}{32}(1 + ci)^2(1 + ck)^2 \cos 2\sigma \\ + \frac{1}{8} sisk(1 + ci)(1 + ck) \cos(2\sigma + \rho) \\ + \frac{3}{16} si^2 sk^2 \cos(2\sigma + 2\rho) \\ + \frac{1}{8} sisk(1 - ci)(1 - ck) \cos(2\sigma + 3\rho) \\ + \frac{1}{32}(1 - ci)^2(1 - ck)^2 \cos(2\sigma + 4\rho) \end{array} \right.$$

$$-(\xi_q^2 - \eta_q^2) \left\{ \begin{array}{l} \frac{1}{8}(2 - 3si^2)(2 - 3sk^2) \\ + \frac{3}{2} siciskck \cos \rho + \frac{3}{8} si^2 sk^2 \cos 2\rho, \end{array} \right. .$$

si and ci stand for $\sin i$ and $\cos i$, while sk and ck stand for $\sin K$ and $\cos K$.

Writing up the differential equations generated by this Hamiltonian, it is easy to find out that they admit the equilibria defined by:

- (A) $\sigma = 0$: the axe of smallest inertia point toward the perturber;
- (B) $\xi_q = \eta_q = 0$: the axe of largest inertia is aligned with the angular momentum;
- (C) $\rho = 0$: the line of node of the orbit and the line of node of the equator are aligned;
- (D) $\frac{\partial \mathcal{H}}{\partial R} = 0$, fixes the value of the obliquity K . It is equivalent to:

$$2\dot{\Omega} \sin K + \frac{n^2 C}{S} \{ [2\delta_1 + \delta_2] \sin(2i - 2K) + 2\delta_2 \sin(i - K) \} = 0 ; \quad (8)$$

- (E) $\frac{\partial \mathcal{H}}{\partial S} = 0$, fixes the value of S . It is equivalent to:

$$S/C = n - \frac{n^2 C}{S} \frac{(1 - \cos K)}{4} \{ [2\delta_1 + \delta_2] \sin(2K - 2i) + 2\delta_2 \sin(K - i) \} .$$

For small value of the inclination i (in the case of the Moon $i = 5^\circ.133$ and $\dot{\Omega} = -4.021 \cdot 10^{-3}$), there exists a solution of equation (8) with small K ; it is given by:

$$K \approx \frac{n^2 C}{S \dot{\Omega}} \frac{-(\delta_1 + \delta_2) i}{[1 - \frac{2n^2 C}{S \dot{\Omega}} (\delta_1 + \delta_2)]} \approx -1^\circ.712 .$$

The full evalation of the formula (8) yields $K = -1^\circ.535$. Of course we could have taken $\rho = \pi$ and then the sign of the obliquity K would have become positive.

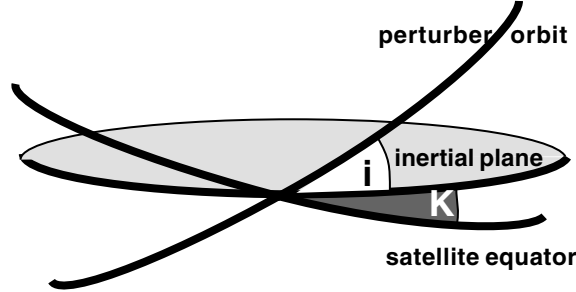


Fig. 2. The second law of Cassini; the nodes of the orbit and of the equator are aligned. For negative value of the precession, the ascending node of the orbit is aligned with the descending node of the equator

5 Motion around the Cassini's States

Dissipative forces due to the non elasticity of the Moon should drive it toward a Cassini's state; but small systematic disturbances (perturbation by other bodies, non sphericity of the Earth, etc.) will induce permanent *forced librations* around the equilibrium. There could be also temporary *free librations* due for instance to collisions with comets. It is thus important to analyze the stability of the Cassini's states and to determine the frequencies of the system linearized around it. A resonance between the frequencies of libration or between one frequency of libration and a driving frequency may result in an important perturbation or even destabilize the equilibrium. In any case, the linearized approximation will provide the integrable problem, stepping stone of a further perturbation step by which the forced long-period librations may be computed. Let us remember that the short period librations should have been computed already while averaging the Hamiltonian on the fast variables (see beginning of Sect. 4).

In order to linearize the differential equations around the equilibrium, we define the *canonical small variations* ξ, η :

$$\begin{aligned} \xi_\sigma &= \sigma & , & & \eta_\sigma &= S - S^* \\ \xi_\rho &= \rho & , & & \eta_\rho &= R - R^* \\ \xi_q &= \xi_q & , & & \eta_q &= \eta_q \end{aligned} \quad (9)$$

where S^* and R^* are the values of S and R at the equilibrium.

Expanding the Hamiltonian function in power of these quantities and keeping only the quadratic terms (the linear terms disappears as the origin is an equilibrium) we find:

$$\begin{aligned} 2\mathcal{H} &= \gamma_{\sigma\sigma}\xi_\sigma^2 + 2\gamma_{\sigma\rho}\xi_\sigma\xi_\rho + \gamma_{\rho\rho}\xi_\rho^2 \\ &\quad + \mu_{\sigma\sigma}\eta_\sigma^2 + 2\mu_{\sigma\rho}\eta_\sigma\eta_\rho + \mu_{\rho\rho}\eta_\rho^2 \\ &\quad + \nu_{\xi\xi}\xi_q^2 + \nu_{\eta\eta}\eta_q^2 . \end{aligned} \quad (10)$$

In the case of the Moon, the value of the coefficients are:

$$\begin{aligned}\gamma_{\sigma\sigma} &= 6.76979 \cdot 10^{-4}, & \gamma_{\sigma\rho} &= -8.136 \cdot 10^{-7}, & \gamma_{\rho\rho} &= -2.6283 \cdot 10^{-6} \\ \mu_{\sigma\sigma} &= 1.00000089, & \mu_{\sigma\rho} &= 1.5505 \cdot 10^{-3}, & \mu_{\rho\rho} &= -4.33224533 \\ \nu_{\xi\xi} &= 2.4728746 \cdot 10^{-3}, & \nu_{\eta\eta} &= 3.913066 \cdot 10^{-4}\end{aligned}\quad (11)$$

The differential equations generated by this quadratic Hamiltonian are linear. They are thus integrable and we can define *angles-actions* variables to describe them in the simplest way. We shall proceed in two steps. First introduce a canonical change of variables to reduce the Hamiltonian to a sum of squares of the variables and then introduce a scaling and polar coordinates in order to define the angles-actions variables.

For the first step we propose the canonical transformation:

$$\begin{aligned}\xi_\sigma &= \zeta_\sigma - \beta\zeta_\rho, & \eta_\sigma &= (1 - \alpha\beta)\chi_\sigma - \alpha\chi_\rho \\ \xi_\rho &= \alpha\zeta_\sigma + (1 - \alpha\beta)\zeta_\rho, & \eta_\rho &= \beta\chi_\sigma + \chi_\rho.\end{aligned}\quad (12)$$

The parameter α and β have to be adjusted in order to eliminates the cross product terms. The coefficient of the term in $\zeta_\sigma\zeta_\rho$ is equal to:

$$\gamma'_{\sigma\rho} = 2\gamma_{\rho\rho}\alpha(1 - \alpha\beta) + 2\gamma_{\sigma\rho}(1 - 2\alpha\beta) - 2\gamma_{\sigma\sigma}\beta = 0, \quad (13)$$

and the coefficient in $\chi_\sigma\chi_\rho$ equal to:

$$\mu'_{\sigma\rho} = 2\mu_{\rho\rho}\beta + 2\mu_{\sigma\rho}(1 - 2\alpha\beta) - 2\mu_{\sigma\sigma}\alpha(1 - \alpha\beta) = 0, \quad (14)$$

Recognizing that the two equations are linear in β , we find:

$$\beta = \frac{\gamma_{\rho\rho}\alpha + \gamma_{\sigma\rho}}{\gamma_{\rho\rho}\alpha^2 + 2\gamma_{\sigma\rho}\alpha + \gamma_{\sigma\sigma}} = \frac{\mu_{\sigma\sigma}\alpha - \mu_{\sigma\rho}}{\mu_{\sigma\sigma}\alpha^2 - 2\mu_{\sigma\rho}\alpha + \mu_{\rho\rho}}, \quad (15)$$

with α root of the quadratic equation which makes the previous equations consistent:

$$A\alpha^2 + B\alpha - C = 0, \quad (16)$$

where

$$A = \gamma_{\sigma\rho}\mu_{\sigma\sigma} + \gamma_{\rho\rho}\mu_{\sigma\sigma}, \quad B = \gamma_{\sigma\sigma}\mu_{\sigma\sigma} - \gamma_{\rho\rho}\mu_{\rho\rho}, \quad C = \gamma_{\sigma\sigma}\mu_{\sigma\rho} + \gamma_{\sigma\rho}\mu_{\rho\rho}.$$

The coefficient A vanishes with δ_2 (if the body has a cylindrical symmetry). In that case, the equation in α has a unique root. In case $\delta_2 \neq 0$, we choose the smallest root which happens to be quite small when the inclination i is small. Correspondingly, the coefficient β is also small. Hence the transformation (12) is close to the identity. In the case of the Moon, we have $\alpha = 1.2285 \cdot 10^{-3}$ and $\beta = -6.8773 \cdot 10^{-3}$, and the Hamiltonian (10) becomes:

$$2\mathcal{H} = [\gamma'_{\sigma\sigma}\zeta_\sigma^2 + \mu'_{\sigma\sigma}\chi_\sigma^2] + [\gamma'_{\rho\rho}\zeta_\rho^2 + \mu'_{\rho\rho}\chi_\rho^2] + [\nu_{\xi\xi}\xi_q^2 + \nu_{\eta\eta}\eta_q^2], \quad (17)$$

with $\gamma'_{\sigma\sigma} = 6.7697924 \cdot 10^{-4}$, $\mu'_{\sigma\sigma} = 0.9999905$, $\gamma'_{\rho\rho} = -2.62937 \cdot 10^{-6}$ and $\mu'_{\rho\rho} = -4.3322926$. The system is the sum of three decoupled linear systems and the angle-action variables are given by:

$$\begin{aligned} \zeta_\sigma &= \sqrt{2UU^*} \sin u \quad , \quad \chi_\sigma = \sqrt{2U/U^*} \cos u \quad , \quad \text{with } U^* = \sqrt{\mu'_{\sigma\sigma}/\gamma'_{\sigma\sigma}} \\ \zeta_\rho &= \sqrt{2VV^*} \sin v \quad , \quad \chi_\rho = \sqrt{2V/V^*} \cos v \quad , \quad \text{with } V^* = \sqrt{\mu'_{\rho\rho}/\gamma'_{\rho\rho}} \\ \xi_q &= \sqrt{2WW^*} \sin w \quad , \quad \eta_q = \sqrt{2W/W^*} \cos w \quad , \quad \text{with } W^* = \sqrt{\nu_{\eta\eta}/\nu_{\xi\xi}} \end{aligned} \quad (18)$$

Finally, after this last transformation, the Hamiltonian is expressed in angles-actions coordinates and read:

$$\mathcal{H} = \omega_u U + \omega_v V + \omega_w W \quad , \quad (19)$$

where the frequencies $(\omega_u, \omega_v, \omega_w)$ are given by:

$$\omega_u = \sqrt{\gamma'_{\sigma\sigma}\mu'_{\sigma\sigma}} \quad , \quad \omega_v = \sqrt{\gamma'_{\rho\rho}\mu'_{\rho\rho}} \quad , \quad \omega_w = \sqrt{\nu_{\xi\xi}\nu_{\eta\eta}} \quad . \quad (20)$$

The period of the three modes of librations are $2\pi/\omega_u = 2.87$ years (free libration in longitude), $2\pi/\omega_v = 22.16$ years (free libration in obliquity) and $2\pi/\omega_w = 76.04$ years (free wobble). A more precise theory based on a more complete model (Moons, 1982) gives $2\pi/\omega_u = 2.88$ years, $2\pi/\omega_v = 24.14$ years and $2\pi/\omega_w = 75.23$ years.

Acknowledgements

The author acknowledges the support of INTAS (00-240)

References

1. Andoyer, H., 1926, *Mécanique Céleste*, Gauthier-Villars, Paris
2. Cassini, G.D.: 1730, De l'origine et du progrès de l'astronomie et de son usage dans la géographie et la navigation, *Mem. Acad. Sc. Paris*, **8**, 1–50
3. Deprit, A.: 1967, Free rotation of an Rigid body studied in the Phase space, *Am. J. of Physics*, **35**, 424–428
4. Goldreich, P. and Peale, S.J.: 1966, Spin-Orbit Coupling in the Solar System, *Astron. J.*, **71**, 425–437
5. Henrard, J. and Murigande, C.: 1987, Colombo's Top, *Celest.l Mech. & Dyn. Astr.*, **40**, 345–366,
6. Henrard, J. and Schwanen, G.: 2004, Rotation of Synchronous Satellites: Application to the Galilean satellites, *Celest.l Mech. & Dyn. Astr.*, submitted
7. Moons, M.: 1982, Physical libration of the Moon, *Celestial Mechanics*, **26**, 131–142
8. Moons, M.: 1984, Planetary Perturbations on the Librations of the Moon, *Celestial Mechanics*, **34**, 263–274
9. Peale, S.J.: 1969, Generalized Cassini's Law, *Astron. J.*, **74**, 483–489

Spin-Orbit Resonant Rotation of Mercury

Sandrine D’Hoedt and Anne Lemaitre

Unité de Systèmes Dynamiques – Département de Mathématique
Facultés Universitaires Notre-Dame de la Paix
Rempart de la Vierge, 8 B 5000 – NAMUR – Belgium
`sandrine.dhoedt@fundp.ac.be`

Abstract. One of the main characteristics of Mercury is its 3 : 2 spin-orbit resonance. The analytical 2-degrees of freedom model proposed here takes into account this phenomenon thanks to the introduction of two suitable resonant variables. Our model must be considered as a base for future computations; it does not include perturbations due to the influence of the other planets, the non-alignment of spin axis with the normal to the equatorial plane and the non-rigidity of the body. These perturbations will be added later on. However the results obtained by our simplified model for the angular variables frequencies are coherent with those given by existing complete numerical models.

1 Introduction

Mercury has received few attention from the astronomical and spatial community; the last and unique space mission concerning Mercury was Mariner 10 launched in 1973. The spacecraft flew by Mercury three times in a retrograde heliocentric orbit and returned images and data of the planet. The scientific objectives of the mission were to measure Mercury’s environment, atmosphere, surface, and body characteristics. Mariner 10 encountered the orbit of Mercury on March 29, 1974, at a distance of about 704 km from the surface, on September 21, 1974, at an altitude of 48,069 km, and on March 16, 1975, at an altitude of 327 km. The mission concluded that Mercury had no atmosphere, a cratered Moon-like surface, a small magnetic field and a relatively large iron-rich core. Concerning the gravitational field coefficients, estimations about J_2 and C_{22} were given with large errors; the most precise values available now are taken from [1] and are : $J_2 = (6.0 \pm 2.0)10^{-5}$ and $C_{22} = (1.0 \pm 0.5)10^{-5}$.

For the other coefficients, the values are still purely speculative.

Two recent space missions have been programmed to explore the planet Mercury : BepiColombo¹, scheduled for a launch in 2012, and MESSENGER², launched on the 3rd August 2004, to *unlock the secrets of Mercury, the last unknown terrestrial planet*. This double departure to the forgotten planet represents a new challenge for building complete or complementary models, numerical as well as analytical, for the rigid body and its inner structure.

The particularity of Mercury, discovered in the sixties, is to be captured in a 3 : 2 spin-orbit resonance; this is the only case in the Solar System, some satellites like the Moon or Europa, are blocked in a 1 : 1 synchronous resonance, but the ratio 3 : 2 for such a resonance is unique. The papers written in the years following this discovery, were mainly devoted to the analysis of mechanisms explaining the capture in such an unusual situation, or to the description of this resonant motion. Let us mention the pioneer papers of [11] and [2, 5, 8, 12] or [4] as examples. The explanation usually accepted is that Mercury was pushed in this resonance by the tidal effects; reasonable values for the probabilities of capture were deduced, to explain the present observed situation, but they did not really privilege this resonance with respect to the other ones, like the 2 : 1 for example. However, recently, with the new numerical techniques and the power of the nowadays computers, [9] published very convincing results, showing that the probability of capture of Mercury in such a 3 : 2 resonance was higher than 0.5, as a result of its chaotic orbital dynamics.

To compute and reproduce the behavior of this insolite motion, simplified models were proposed by the authors mentioned above, as toys models, to describe the effect of the dissipations on the capture process. Unfortunately, probably due to the lack of data and the complexity of the resonant rotation, complete analytical models of rotation, using classical perturbation techniques (developed for other planets or satellites), were not applied for Mercury.

The interest of getting such a model for the space missions is obvious, even if, besides, very sophisticated numerical integrations are performed. The analytical model of rotation can be used to analyze the effect of each contribution on the motion of the satellite, through the corresponding variation equations. Let us mention, as examples, the coefficients of the gravitational field, the obliquity or the combined motion of the axes of figure and of the spin axis. The analytical methods allow to keep the unknown constants (like the higher zonal and tesseral coefficients) as parameters and to stress their influence on several variables.

Furthermore analytical orbital theories are available for Mercury, and some of them can give a precise motion over periods of time of 100 years; the idea is to build a model of rotation quite complete, in which a theory of motion is

¹Mission of the European Space Agency and ISAS, Japan’s Institute of Space Astronautical Sciences

²Mercury Surface Space ENvironment, Geochemistry and Ranging, spatial mission of NASA

injected as a black box. This idea has been applied to the satellite of Jupiter, Europa, by [7] using the theory of [10] with success.

For Mercury the initial step is to use a Keplerian orbit to get a first model and to complete it through updated versions of VSOP82 [3], available in the IMCCE in Paris.

Our future ambition is to build a complete model of rotation of Mercury, based on a precise orbital theory, in Hamiltonian formalism, using angle-action variables and giving, at the end, the three main frequencies of the rotation (and consequently the motion of the three Euler angles). These frequencies are first evaluated in a very simple and basic model, which we present here, and will be later adjusted by adding successive perturbations to the model. The process is based on Lie triangles up to (probably) the second or third order.

This paper presents the kernel, in the form of a two degrees of freedom Hamiltonian system, averaged over the short periods; it succeeds in describing the libration about the 3 : 2 resonant spin-orbit motion by two pairs of action-angle variables, obtained after a succession of simplifications and canonical changes of variables. The two basic frequencies of the system are in complete agreement with the very recent values obtained numerically by [13]

The basic hypotheses are the following : Mercury is a rigid non spherical body, the gravity field is truncated after the second degree terms (the only ones for which we have significant numerical values), the orbital motion of Mercury is Keplerian and the spin axis is parallel to the third axis of inertia. Let us remark that the obliquity is not put to zero (the spin axis is not perpendicular neither to the orbital plane nor to the ecliptic plane).

2 Reference Frames and Variables Choice

We are going to work with four reference frames all centered at Mercury's center of mass.

The three first frames are:

- the inertial frame (X_0, Y_0, Z_0) linked to the Ecliptic,
- the Spin frame (X_2, Y_2, Z_2) with Z_2 in the direction of the angular momentum and
- the Body frame (X_3, Y_3, Z_3) with Z_3 in the direction of the axis of smallest inertia.

These frames are linked together by two sets of Euler's angles: $(h, K, -)$ and (g, J, l) .

The fourth frame (X_1, Y_1, Z_1) is linked to the keplerian orbital plane of Mercury with Ω_o the longitude of the ascending node, ω_o the argument of the pericenter and i_o the inclination.

Let us choose as first set of variables the Delaunay's elements associated to the elliptic elements of Mercury's center of mass:

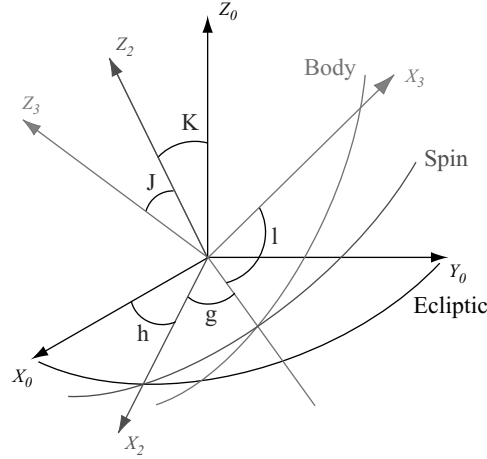


Fig. 1. The three frames (X_0, Y_0, Z_0) , (X_2, Y_2, Z_2) and (X_3, Y_3, Z_3) , all centered at Mercury's center of mass with $(h, K, \cdot)$ being Euler's angles related to (X_0, Y_0, Z_0) and (X_2, Y_2, Z_2) and (g, J, l) being Euler's angles related to (X_2, Y_2, Z_2) and (X_3, Y_3, Z_3)

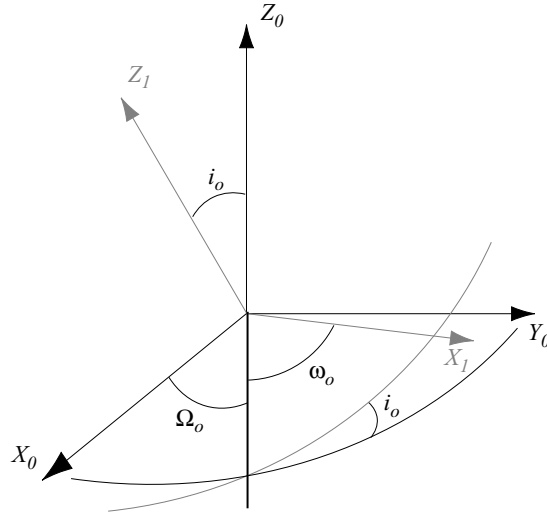


Fig. 2. The orbital frame (X_1, Y_1, Z_1) , centered at Mercury's center of mass with $(\Omega_o, i_o, \omega_o)$ being Euler's angles related to (X_0, Y_0, Z_0) and (X_1, Y_1, Z_1)

$$L_o, G_o, H_o, l_o, g_o, h_o$$

with l_o the mean anomaly, $g_o = \omega_o$, $h_o = \Omega_o$, $L_o = m\sqrt{\mu a}$, $G_o = L_o\sqrt{1-e^2}$ and $H_o = G_o \cos i_o$, with a the semi major axis, e the eccentricity and m the mass of Mercury.

Andoyer's variables complete the set of canonical variables:

$$L, G, H, l, g, h$$

with G , the angular momentum (in the direction of Z_2), $L = G \cos J$ (projection of the angular momentum on Z_3), $H = G \cos K$ (projection of the angular momentum on Z_0).

Let us note that if J (resp. K) is zero, l and g (resp. g and h) are no more well defined but their sum is always.

In order to avoid these definition problems, we are going to perform a canonical transformation to have new partially non singular variables (λ_1 is always well defined)

$$\begin{aligned}\lambda_1 &= l + g + h \\ \lambda_2 &= -l \\ \lambda_3 &= -h.\end{aligned}$$

Their conjugated momenta are given by:

$$\begin{aligned}\Lambda_1 &= G \\ \Lambda_2 &= G - L = G(1 - \cos J) \\ \Lambda_3 &= G - H = G(1 - \cos K).\end{aligned}$$

3 First Model of Rotation

Without planetary perturbations, the Hamiltonian can be written as:

$$\mathcal{H} = -\frac{\mu^2}{2L_o^2} + T(\Lambda_1, \Lambda_2, \Lambda_3, \lambda_1, \lambda_2, \lambda_3) + V_G$$

where the kinetic energy of rotation T is

$$T = \frac{(\Lambda_1 - \Lambda_2)^2}{2I_3} + \frac{1}{2}(\Lambda_1^2 - (\Lambda_1 - \Lambda_2)^2) \left(\frac{\sin^2 \lambda_2}{I_1} + \frac{\cos^2 \lambda_2}{I_2} \right)$$

with I_1, I_2, I_3 the inertia momenta ($I_1 < I_2 < I_3$) and the potential V_G due to the gravitational attraction between the two bodies is given by

$$V_G = -\frac{GMm}{r} \left(\frac{R_e}{r} \right)^2 [C_2^0 P_2(\sin \theta) + C_2^2 P_2^2(\sin \theta) \cos 2\varphi]$$

with M and m the masses of the Sun and Mercury respectively, R_e the equatorial radius of Mercury, r the distance between Mercury and Sun centers of mass, θ and φ the latitude and the longitude in (X_3, Y_3, Z_3) and where the development in Legendre's polynomials has been limited to the 2nd-order in harmonics.

4 Development of the Gravitational Potential

Let us now express V_G in terms of cartesian coordinates:

$$\begin{cases} \bar{x}_3 = \cos \varphi \cos \theta \\ \bar{y}_3 = \sin \varphi \cos \theta \\ \bar{z}_3 = \sin \theta \end{cases}$$

where $(\bar{x}_3, \bar{y}_3, \bar{z}_3)$ is the unit vector in the direction of the perturbing body (the subscript “3” being used to recall that we are working in the Body frame (X_3, Y_3, Z_3)).

Replacing Legendre's polynomials by their expression, we find:

$$V_G = -\frac{GMm}{r^3} R_e^2 \left[\frac{C_2^0}{2} (2\bar{z}_3^2 - \bar{x}_3^2 - \bar{y}_3^2) + 3C_2^2 (\bar{x}_3^2 - \bar{y}_3^2) \right].$$

We have now to express the unit vector in the frame linked to the orbit (X_1, Y_1, Z_1) :

$$\begin{pmatrix} \bar{x}_3 \\ \bar{y}_3 \\ \bar{z}_3 \end{pmatrix} = M_3(l)M_1(J)M_3(g)M_1(K)M_3(h) \\ \times M_3(-h_o)M_1(-i_o)M_3(-\omega_o) \begin{pmatrix} \cos v_o \\ \sin v_o \\ 0 \end{pmatrix}.$$

From now on, we are going to make the following approximation:

$$J = 0.$$

The choice is often made in the literature (e.g. [11]) because it seems to fit very well the reality. Thus in our variables:

$$\begin{pmatrix} \bar{x}_3 \\ \bar{y}_3 \\ \bar{z}_3 \end{pmatrix} = M_3(\lambda_1 + \lambda_3)M_1(K)M_3(-\lambda_3)M_3(-h_o)M_1(-i_o)M_3(-g_o) \begin{pmatrix} \cos v_o \\ \sin v_o \\ 0 \end{pmatrix}.$$

V_G can now be written like this:

$$V_G = -\frac{GMm}{a^3} R_e^2 \left(\frac{a}{r} \right)^3 (k_c \cos^2 v_o + k_{cs} \cos v_o \sin v_o + k_s \sin^2 v_o) \quad (1)$$

where k_c , k_{cs} and k_s are function of $(\lambda_1, \lambda_3, K, h_o, i_o, g_o)$.

We want to make appear the mean anomaly l_o and its conjugated momentum L_o in the expression of V_G . The introduction of l_o in (1) can be done thanks to classical developments (here limited to the 3rd order in eccentricity):

$$\begin{aligned}
\frac{r}{a} &= 1 + \frac{1}{2} e^2 + \left(-e + \frac{3}{8} e^3\right) \cos l_o - \frac{1}{2} e^2 \cos 2l_o - \frac{3}{8} e^3 \cos 3l_o \\
\frac{r}{a} \cos v_o &= -\frac{3}{2} e + \left(1 - \frac{3}{8} e^2\right) \cos l_o + \left(\frac{1}{2} e - \frac{1}{3} e^3\right) \cos 2l_o \\
&\quad + \frac{3}{8} e^2 \cos 3l_o + \frac{1}{3} e^3 \cos 4l_o \\
\frac{r}{a} \sin v_o &= \left(1 - \frac{5}{8} e^2\right) \sin l_o + \left(\frac{1}{2} e - \frac{5}{12} e^3\right) \sin 2l_o \\
&\quad + \frac{3}{8} e^2 \sin 3l_o + \frac{1}{3} e^3 \sin 4l_o
\end{aligned}$$

and, consequently, with $L_o = m\sqrt{\mu a}$,

$$\begin{aligned}
V_G &= -\frac{GMm^7}{L_o^6} \mu^3 R_e^2 \times \\
&\quad \left\{ \frac{1}{2} C_2^0 \left[a_{000} + \sum_{i=1}^5 a_{00i} \cos(i l_o) + \sum_{j=1}^2 \sum_{i=-5}^5 a_{0ji} \cos(j \lambda_3 + i l_o) \right. \right. \\
&\quad \left. \left. + \sum_{i=1}^5 b_{00i} \sin(i l_o) + \sum_{j=1}^2 \sum_{i=-5}^5 b_{0ji} \sin(j \lambda_3 + i l_o) \right] \right. \\
&\quad \left. + 3C_2^2 \left[\sum_{i=-5}^5 a_{20i} \cos(2\lambda_1 + i l_o) + \sum_{j=1}^4 \sum_{i=-5}^5 a_{2ji} \cos(2\lambda_1 + j \lambda_3 + i l_o) \right. \right. \\
&\quad \left. \left. + \sum_{i=-5}^5 b_{20i} \sin(2\lambda_1 + i l_o) + \sum_{j=1}^4 \sum_{i=-5}^5 b_{2ji} \sin(2\lambda_1 + j \lambda_3 + i l_o) \right] \right\}
\end{aligned}$$

where the coefficients a_{ijk} and b_{ijk} depend on K, h_o, i_o, g_o and e , with h_o, i_o, g_o and e considered here as constants (as long as we neglect the planetary perturbations). So, in our basic model, V_G is only function of 3 momenta and 3 angles : Λ_1 and Λ_3 (through K), L_o , λ_1, λ_3 and l_o .

5 Spin-Orbit Resonant Angle

Let us recall that Mercury is captured in a 3 : 2 spin-orbit resonance which can be translated in our variables by:

$$\dot{\lambda}_1 = \frac{3}{2} \dot{l}_o.$$

Let us thus take

$$\sigma = \frac{2\lambda_1 - 3l_o}{2}$$

as the resonant angle and keep Λ_1 as its conjugated momentum. To keep a canonical transformation, we have to associate to l_o a new momentum $\Lambda_0 = L_o + \frac{3}{2} \Lambda_1$. The potential is then given by:

$$\begin{aligned}
V_G = & -\frac{GMm^7}{(\Lambda_0 - \frac{3}{2}\Lambda_1)^6} \mu^3 R_e^2 \\
& \left(\frac{1}{2} C_2^0 (a_{000} + \sum_{k=1}^2 (a_{0k0} \cos(k\lambda_3) + b_{0k0} \sin(k\lambda_3)) \right. \\
& \quad + \sum_{i=1}^5 (a_{00i} \cos(i l_0) + b_{00i} \sin(i l_0)) \\
& \quad + \sum_{i=-5, i \neq 0}^5 (a_{01i} \cos(\lambda_3 + i l_0) + b_{01i} \sin(\lambda_3 + i l_0)) \\
& \quad + \sum_{i=-5, i \neq 0}^5 (a_{02i} \cos(2\lambda_3 + i l_0) + b_{02i} \sin(2\lambda_3 + i l_0)) \\
& \quad + 3 C_2^2 (a_{200} \cos 2\sigma + b_{200} \sin 2\sigma + \sum_{k=1}^4 (a_{2k0} \cos(2\sigma + k\lambda_3) \\
& \quad \quad \quad + b_{2k0} \sin(2\sigma + k\lambda_3)) \\
& \quad + \sum_{j=-2}^8 (a_{20j} \cos(2\sigma + j l_0) + b_{20j} \sin(2\sigma + j l_0)) \\
& \quad + \sum_{k=1}^4 \sum_{j=-2}^8 (a_{2kj} \cos(2\sigma + k\lambda_3 + j l_0) \\
& \quad \quad \quad + b_{2kj} \sin(2\sigma + k\lambda_3 + j l_0)) \Big)
\end{aligned}$$

with i and j not equal to zero. One can see that the term in C_2^2 is the only one containing the resonant angle.

Let us now average the potential on l_o ; $\langle V_G \rangle$ is given by the following expression, where $\langle \rangle$ stands for "averaged".

$$\langle V_G \rangle = -\frac{GMm^7}{(\Lambda_0 - \frac{3}{2}\Lambda_1)^6} \mu^3 R_e^2 \left(\frac{1}{2} C_2^0 \gamma_1 + 3 C_2^2 \gamma_2 \right)$$

where

$$\begin{aligned}
\gamma_1 = & \left(1 + \frac{3e^2}{2} \right) \left[-\frac{1}{4} \left(-1 + 3 \cos^2 i_o \right) \left(-1 + 3 \cos^2 K \right) \right. \\
& - 3 \cos i_o \cos K \cos(h_o + \lambda_3) \sin i_o \sin K \\
& \left. - \frac{3}{4} \left(1 - \cos^2 i_o \right) \left(1 - \cos^2 K \right) \cos(2h_o + 2\lambda_3) \right] \\
\gamma_2 = & \frac{1}{16} \left(\frac{7e}{2} - \frac{123e^3}{16} \right) \left(1 - \cos i_o \right)^2 \left(1 - \cos K \right)^2 \cos(2h_o + 4\lambda_3 + 2\sigma - 2g_o) \\
& + \frac{53e^3}{128} \left(1 - \cos K \right)^2 \sin^2 i_o \cos(2h_o + 4\lambda_3 + 2\sigma) \\
& + \frac{1}{4} \left(\frac{7e}{2} - \frac{123e^3}{16} \right) \left(1 - \cos i_o \right) \left(1 - \cos K \right) \sin i_o \\
& \quad \sin K \cos(h_o + 3\lambda_3 + 2\sigma - 2g_o)
\end{aligned}$$

$$\begin{aligned}
& + \frac{53e^3}{32} \cos i_o \left(1 - \cos K\right) \sin i_o \sin K \cos(h_o + 3\lambda_3 + 2\sigma) \\
& + \frac{3}{8} \left(\frac{7e}{2} - \frac{123e^3}{16}\right) \sin^2 i_o \sin^2 K \cos(2\lambda_3 + 2\sigma - 2g_o) \\
& + \frac{53e^3}{64} \left(2 - 3 \sin^2 i_o\right) \sin^2 K \cos(2\lambda_3 + 2\sigma) \\
& + \frac{1}{4} \left(\frac{7e}{2} - \frac{123e^3}{16}\right) \left(1 + \cos i_o\right) \left(1 + \cos K\right) \sin i_o \\
& \quad \sin K \cos(h_o - \lambda_3 - 2\sigma + 2g_o) \\
& - \frac{53e^3}{32} \cos i_o \left(1 + \cos K\right) \sin i_o \sin K \cos(h_o - \lambda_3 - 2\sigma) \\
& + \frac{1}{16} \left(\frac{7e}{2} - \frac{123e^3}{16}\right) \left(1 + \cos i_o\right)^2 \left(1 + \cos K\right)^2 \cos(2h_o - 2\sigma + 2g_o) \\
& + \frac{53e^3}{128} \left(1 + \cos K\right)^2 \sin^2 i_o \cos(2h_o - 2\sigma)
\end{aligned}$$

6 Simplified Hamiltonian and Basic Frequencies

Considering that the eccentricity $e \sim 0.206$, we are going to keep the terms in $(1 + \frac{3e^2}{2})$ and $(\frac{7e}{2} - \frac{123e^3}{16})$ but not in e^3 only.

And so, we obtain a two degrees of freedom Hamiltonian:

$$\begin{aligned}
\langle \mathcal{H} \rangle = & \frac{\Lambda_1^2}{2I_3} - \frac{m^3 \mu^2}{2(\Lambda_0 - \frac{3\Lambda_1}{2})^2} - \frac{GMm^7 \mu^3 R_e^2}{(\Lambda_0 - \frac{3\Lambda_1}{2})^6} \\
& \times \left[\frac{1}{2} C_2^0 \left(1 + \frac{3e^2}{2}\right) \left(-\frac{1}{4} \left(-1 + 3 \cos^2 i_o\right) \left(-1 + 3 \cos^2 K\right) \right. \right. \\
& - 3 \cos i_o \cos K \cos(\sigma_3) \sin i_o \sin K \\
& - \frac{3}{4} \left(1 - \cos^2 i_o\right) \left(1 - \cos^2 K\right) \cos(2\sigma_3) \Big) \\
& + 3 C_2^2 \left(\frac{7e}{2} - \frac{123e^3}{16}\right) \left(\frac{1}{16} \left(1 - \cos i_o\right)^2 \left(1 - \cos K\right)^2 \cos(2\sigma_1 + 4\sigma_3) \right. \\
& + \frac{1}{4} \left(1 - \cos i_o\right) \left(1 - \cos K\right) \sin i_o \sin K \cos(2\sigma_1 + 3\sigma_3) \\
& + \frac{3}{8} \sin^2 i_o \sin^2 K \cos(2\sigma_1 + 2\sigma_3) \\
& + \frac{1}{4} \left(1 + \cos i_o\right) \left(1 + \cos K\right) \sin i_o \sin K \cos(2\sigma_1 + \sigma_3) \\
& \left. \left. + \frac{1}{16} \left(1 + \cos i_o\right)^2 \left(1 + \cos K\right)^2 \cos(2\sigma_1) \right) \right]
\end{aligned}$$

with

$$\sigma_1 = \sigma - h_o - g_o, \quad \sigma_3 = \lambda_3 + h_o$$

and keeping Λ_1 and Λ_3 as their conjugated momenta.

In order to compute the equilibria, we have to put to zero the right-hand sides of the differential equations, where we consider $\langle \mathcal{H} \rangle$ as a function of $\cos K$, Λ_1 , σ_1 and σ_3 , in which $\cos K$ is dependent on Λ_1 and Λ_3 .

$$\begin{aligned} \dot{\sigma}_1 &= \frac{\partial \langle \mathcal{H} \rangle}{\partial \Lambda_1} + \frac{\partial \langle \mathcal{H} \rangle}{\partial (\cos K)} \frac{\partial (\cos K)}{\partial \Lambda_1} \\ \dot{\sigma}_3 &= \frac{\partial \langle \mathcal{H} \rangle}{\partial (\cos K)} \frac{\partial (\cos K)}{\partial \Lambda_3} \\ \dot{\Lambda}_1 &= -\frac{\partial \langle \mathcal{H} \rangle}{\partial \sigma_1} = \sum_{i=0}^4 d_i \sin(2\sigma_1 + i\sigma_3) \\ \dot{\Lambda}_3 &= -\frac{\partial \langle \mathcal{H} \rangle}{\partial \sigma_3} = \sum_{i=1}^2 d'_i \sin(i\sigma_3) + \sum_{i=1}^4 d''_i \sin(2\sigma_1 + i\sigma_3), \end{aligned}$$

where the coefficients d_i , d'_i and d''_i are depending on i_o , K and e .

The first two equations lead to:

$$\frac{\partial \langle \mathcal{H} \rangle}{\partial \Lambda_1} = 0 = \frac{\partial \langle \mathcal{H} \rangle}{\partial (\cos K)}$$

and the last two equations furnish four couples of angle equilibria:

$$(\sigma_1, \sigma_3) = (0, 0), (0, \pi), (\pi/2, 0), (\pi/2, \pi)$$

After solving the other two equations, we find 16 equilibria (4 couples of equilibria for (Λ_1, Λ_3) for each couple (σ_1, σ_3)).

To find the stability of these equilibria, we have to substitute all these values in the four variational equations, where $\langle \mathcal{H} \rangle$ is classically considered here as a function of the two momenta Λ_1 and Λ_3 and the two conjugated variables σ_1 and σ_3 :

$$\begin{aligned} d\dot{\sigma}_1 &= \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \sigma_1 \partial \Lambda_1} d\sigma_1 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial^2 \Lambda_1} d\Lambda_1 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \sigma_3 \partial \Lambda_1} d\sigma_3 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \Lambda_3 \partial \Lambda_1} d\Lambda_3 \\ d\dot{\sigma}_3 &= \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \sigma_1 \partial \Lambda_3} d\sigma_1 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \Lambda_1 \partial \Lambda_3} d\Lambda_1 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \sigma_3 \partial \Lambda_3} d\sigma_3 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial^2 \Lambda_3} d\Lambda_3 \\ d\dot{\Lambda}_1 &= -\left(\frac{\partial^2 \langle \mathcal{H} \rangle}{\partial^2 \sigma_1} d\sigma_1 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \Lambda_1 \partial \sigma_1} d\Lambda_1 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \sigma_3 \partial \sigma_1} d\sigma_3 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \Lambda_3 \partial \sigma_1} d\Lambda_3 \right) \\ d\dot{\Lambda}_3 &= -\left(\frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \sigma_1 \partial \sigma_3} d\sigma_1 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \Lambda_1 \partial \sigma_3} d\Lambda_1 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial^2 \sigma_3} d\sigma_3 + \frac{\partial^2 \langle \mathcal{H} \rangle}{\partial \Lambda_3 \partial \sigma_3} d\Lambda_3 \right), \end{aligned}$$

The computation of the eigen values leads to the conclusion that the equilibrium ($\sigma_1 = 0 = \sigma_3$), associated with $K = i_o \sim 7^\circ$, is stable.

The numerical values of the associated momenta (expressed in $\frac{m R_e^2}{year}$) are:

$$\Lambda_1 = 13.30299 \quad \text{and} \quad \Lambda_3 = 0.0991584.$$

Let us remark that this obliquity is measured with respect to the normal to the ecliptic; this equilibrium is the same as in [12], but measured in a different frame. More details about these equilibria and their stability are given in [6].

Mercury being not blocked at the exact 3 : 2 resonance, there is a small libration about this equilibrium. To measure the amplitude and the period of this libration, we have to make a translation to the equilibrium:

$$\begin{aligned} \sigma_1 &= \eta_1 \\ \sigma_3 &= \eta_3 \\ \Lambda_1 &= \Lambda_{10} + \xi_1 \\ \Lambda_3 &= \Lambda_{30} + \xi_3 \end{aligned}$$

Let us now develop the Hamiltonian in series of η_1 , η_3 , ξ_1 and ξ_3 up to order 2, we obtain an Hamiltonian with mixed terms:

$$< \mathcal{H} > = a \eta_1^2 + b \eta_1 \eta_3 + c \eta_3^2 + d \xi_1^2 - e \xi_1 \xi_3 + f \xi_3^2$$

In order to eliminate these mixed terms, we perform again a form-like canonical transformation:

$$\begin{aligned} \eta_1 &= r - \alpha s \\ \eta_3 &= \beta r + s \\ \xi_1 &= R - \gamma S \\ \xi_3 &= \delta R + S. \end{aligned}$$

and we obtain:

$$< \mathcal{H} > = a' r^2 + b' R^2 + c' s^2 + d' S^2 .$$

We still perform a canonical transformation:

$$\begin{aligned} r &= \alpha' u ; & R &= \frac{U}{\alpha'} \\ s &= \beta' v ; & S &= \frac{V}{\beta'} . \end{aligned}$$

the result is:

$$< \mathcal{H} > = A(u^2 + U^2) + B(v^2 + V^2)$$

We have to perform a last canonical transformation to introduce the angle-action coordinates

$$u = \sqrt{2Y} \cos \psi ; \quad U = \sqrt{2Y} \sin \psi$$

$$v = \sqrt{2Z} \cos \zeta ; \quad V = \sqrt{2Z} \sin \zeta$$

and get an Hamiltonian depending only on the moments:

$$\langle \mathcal{H} \rangle = \nu_{\sigma_1} Y^2 + \nu_{\sigma_3} Z^2$$

where $\nu_{\sigma_1} = 2A$ and $\nu_{\sigma_3} = 2B$ are the frequencies of the angles σ_1 and σ_3 .

The periods found by our model are: 15.8573 years for σ_1 and 1065.08 years for σ_3 . These numerical values are in agreement with complete numerical approach of [13].

7 Conclusion

Our kernel model being in agreement with complete numerical models, it will be the base of our further computations: the introduction of the angle J and of the planetary perturbations.

The real purpose of this analytical approach of the rotation of Mercury is the identification of all the frequencies (proper or forced by the planetary motions) playing a significative role in the motion. The future space missions (MESSENGER or BepiColombo) should detect those basic frequencies in the observed motion of the satellite, with slight modifications. If the rotation model is realistic enough, these differences between theory and observations (variational equations) will lead to corrections (or better knowledge) of the parameters of Mercury, like the principal moments of inertia, the coefficients of the gravitational field and the non conservative forces. It could help checking hypotheses about the interior of the planet, and answer questions like the existence and the size of a viscous liquid core.

The challenge, for the scientists, is to provide a sufficiently precise model, so that it could be directly used by the spatial community for comparisons with observations and measurements.

References

1. J.D. Anderson, G. Colombo, P.B. Esposito, E.L. Lau, G.B. and Trager, G.B, *Icarus*, **71**, 337 (1987)
2. V.V. Beletskii, *Celestial Mechanics*, **6**, 359 (1972)
3. P. Bretagnon, *Astronomy and Astrophysics*, **114**, 278 (1982)
4. T.J. Burns, *Celestial Mechanics*, **19**, 297 (1979)
5. G. Colombo, *Nature*, **208**, 575 (1965)
6. S. D'Hoedt and A. Lemaitre, *Celestial Mechanics and Dynamical Astronomy*, **89**, 267 (2004)
7. J. Henrard and G. Schwanen, *Celestial Mechanics and Dynamical Astronomy*, **89**, 181 (2004)

8. H. Kinoshita, Publ. Astron. Soc. Japan, **24**, 423 (1972)
9. A.C. Correia and J. Laskar, Nature, **429**, 848 (2004)
10. V. Lainey, L. Duriez and A. Vienne, Astronomy and Astrophysics, **420**, 1171 (2004)
11. S.J Peale: The Astronomical Journal, **74**, 483 (1969)
12. S.J Peale: The Astronomical Journal, **79**, 722 (1974)
13. N. Rambaux and E. Bois: Astronomy and Astrophysics, **413**, 381 (2004)

Dynamics of Planetary Rings

Bruno Sicardy^{1,2}

¹ Observatoire de Paris, LESIA, 92195 Meudon Cédex, France
bruno.sicardy@obspm.fr

² Université Pierre et Marie Curie, 75252 Paris Cédex 05, France

Abstract. Planetary rings are found around all four giant planets of our solar system. These collisional and highly flattened disks exhibit a whole wealth of physical processes involving microscopic dust grains, as well as meter-sized boulders. These processes, together with ring composition, can help to understand better the formation and evolution of proto-satellite and proto-planetary disks in the early solar system. The present chapter reviews some fundamental aspects of ring dynamics, namely their flattening, their stability against proper modes, their particle sizes, and their responses to resonance forcing by satellites. These concepts will be used and tested during the forthcoming exploration of the Saturn system by the *Cassini* mission.

1 Introduction

Planetary rings consist in thin disks of innumerable colliding particles revolving around a central planet. They are found around all the four giant planets of our solar system, Jupiter, Saturn, Uranus and Neptune. Meanwhile, they exhibit a wide variety of sizes, masses and physical properties. Also, they involve many different physical processes, ranging from large spiral waves akin to galactic structures, down to microscopic electromagnetic forces on charged dust grains.

All these effects have a profound influence on the long term evolution of rings. As such, they can teach us something about the origin of rings, i.e. whether they are cogenetic to the central planet, or the result of a more recent breakup of a comet or satellite. These processes are also linked to disk dynamics in general, either in proto-planetary circumstellar disks or in galaxies.

A complete review of all these processes remains out of the scope of this short chapter. Instead, we would like to address here a few basic issues related to planetary rings, by asking the following questions:

- Why are planetary rings so flat?

- How thin and dynamically stable are they?
- How do they respond to resonant forcing from satellites?

Answering these questions may help to understand the connection between rings global properties (mass, optical depth, etc. . .) and their local characteristics (particle size and distribution, velocity dispersion, etc. . .).

Extensive descriptions and studies of the issues presented in this chapter are available in the literature. Detailed reviews (and related references) on ring structures, ring dynamics and open issues are presented in [1–6], while ring origin and evolution are discussed in [7] and [8]. Stability issues are discussed in classical papers like [9], and in very much details in reference text books like [10]. Collisional processes are reviewed for instance in [11]. Finally, the various possible responses of disks to resonances and their applications are exposed in [1, 12–14].

2 Planetary Rings and the Roche Zone

Generally speaking, planetary rings reside inside a limit loosely referred to as the ‘Roche zone’ of the central planet. Inside this limit, the tidal stress from the central body tends to disrupt satellites into smaller bodies. Outside this limit, accretion tends to sweep dispersed particles into larger lumps.

Reality is not so sharply defined, though. For instance, the tidal disruption limit for a given body depends not only on the bulk density of that body, which may differ significantly from one body to the other, but also on the tensile strength of that body, another parameter which may vary by several orders of magnitude. Thus, due to these uncertainties, the Roche zone depicted in Fig. 1 remains rather blurred.

In spite of these difficulties, it is instructive to plot the sizes of, say, Saturn’s satellites as a function of their distances to the center of the planet (Fig. 1). Interestingly, one notes that the sizes of the satellites decrease as one approaches the Roche zone, as expected from a simple modelling of tidal stress.

Figure 1 also plots the sizes of satellites that could be formed by lumping together the material of Saturn’s A, B and C rings, as well as of Cassini Division. This figure supports the general idea that the Roche Limit delineates the region inside which tides disrupt solid satellites into a fluid-like ring disk.

Note finally that in the Roche zone, rings and satellites can co-exist, as it is the case for instance for Pan, which orbits inside the Encke Division of Saturn’s A ring. An interesting question is then to know whether satellites and rings can influence each other gravitationally, or even if they can be transformed into each other on long (billions of years) or short (a few years) time scales.

Although Jupiter’s, Uranus’ and Neptune’s rings are much less massive than Saturn’s, they exhibit the same general behavior: smaller and smaller

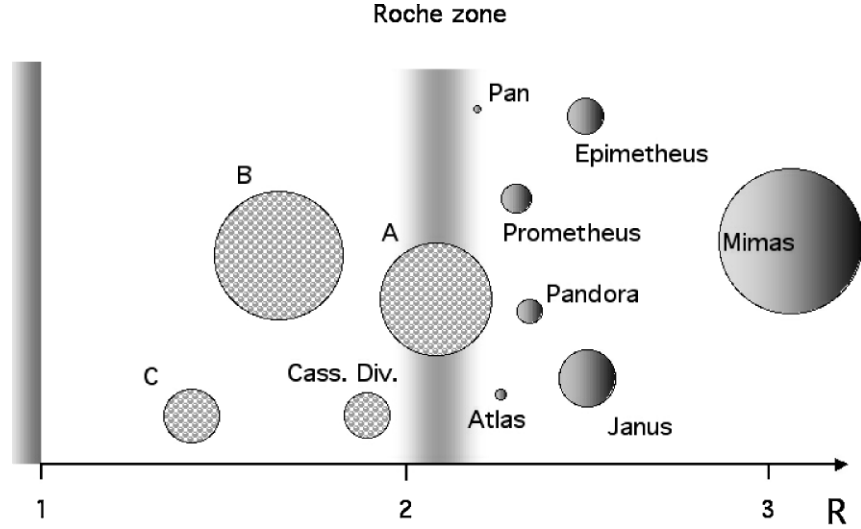


Fig. 1. The relative sizes of Saturn's inner satellites as a function of their distance to the planet center, in units of the planet radius. The "sizes" of the rings have been calculated by lumping all the material of A, B, C rings and Cassini Division into single bodies. All the sizes have been plotted so that to respect the relative masses of the various bodies involved. For comparison, Mimas has a diameter of about 500 km

satellites are encountered as one approaches the planet Roche zone, then a mixture of rings and satellites is observed, and then only rings are found.

3 Flattening of Rings

Planetary rings flatten because of collisions between particles. This process dissipates mechanical energy, while conserving angular momentum of the ensemble.

Consider a swarm of particles labelled $1, \dots, i, \dots$, with total initial angular momentum $\mathbf{H}$ orbiting a spherical planet centered on O . Let us call the Oz axis directed along $\mathbf{H}$ the "vertical" axis, and the plane Oxy perpendicular to Oz the "horizontal plane". The orthogonal unit vectors along Ox , Oy and Oz will be denoted $\hat{\mathbf{x}}$, $\hat{\mathbf{y}}$ and $\hat{\mathbf{z}}$, respectively.

Each particle of mass m_i , position $\mathbf{r}_i$ and velocity $\mathbf{v}_i$ has an angular momentum $\mathbf{H}_i = m_i \mathbf{r}_i \times \mathbf{v}_i$, so that:

$$\mathbf{H} = \sum_i m_i \mathbf{r}_i \times \mathbf{v}_i = \sum_i m_i \mathbf{r}_i \times \mathbf{v}_{iz} + \sum_i m_i \mathbf{r}_i \times \mathbf{v}_{ih} ,$$

where $\mathbf{v}_{iz}$ and $\mathbf{v}_{ih}$ are the vertical and horizontal contributions of the velocities, respectively. The vector $\mathbf{v}_{ih}$ can furthermore be decomposed into a

tangential component, $\mathbf{v}_{i\theta}$ and a radial one, $\mathbf{v}_{ir}$. Projecting $\mathbf{H}$ along the unit vector $\hat{\mathbf{z}}$ along Oz , one gets:

$$H = \mathbf{H} \cdot \hat{\mathbf{z}} = \sum_i m_i (\mathbf{r}_i \times \mathbf{v}_{iz}) \cdot \hat{\mathbf{z}} + \sum_i m_i (\mathbf{r}_i \times \mathbf{v}_{ih}) \cdot \hat{\mathbf{z}} = \sum_i m_i r_i v_{i\theta} .$$

The equality above holds because the first sum vanishes altogether since $\mathbf{v}_{iz}$ and $\hat{\mathbf{z}}$ are parallel, and also because only the tangential velocity survives in the second sum, due to the presence of $\mathbf{r}_i$ in the mixed product $(\mathbf{r}_i \times \mathbf{v}_{ih}) \cdot \hat{\mathbf{z}}$.

On the other hand, the mechanical energy of the ring is given by:

$$E = \sum_i m_i \Phi_P(r_i) + \sum_i m_i v_{i\theta}^2/2 + \sum_i m_i v_{ir}^2/2 + \sum_i m_i v_{iz}^2/2 ,$$

where $\Phi_P(r_i)$ is the planetary potential well per unit mass felt by the particle i .

Comparing the expressions of H and E , one sees that it is possible to minimize the energy E of the ring while conserving the angular momentum H , by simply zeroing v_{ir} and v_{iz} . In other words, the collisions, by dissipating energy while conserving angular momentum, tend to flatten the disk perpendicular to its total angular momentum, and to circularize the particles orbits.

If the planet is not spherically symmetric, as it is the case for all the oblate giant planets, then only the projection of $\mathbf{H}$ along the planet spin axis (say, $\mathbf{H}_{\text{spin}}$) is conserved. The same reasoning as above then shows that the configuration of least energy which conserves $\mathbf{H}_{\text{spin}}$ is a flat, but now equatorial ring.

4 Stability of Flat Disks

The considerations presented above lead to the conclusion that a dissipative ring will eventually collapse to an infinitely thin disk with perfectly circular orbits.

In reality, this is not quite true, as some physics is still missing in the description of the rings. For instance, the differential Keplerian motion, combined with the finite size of the particles and the mutual gravitational stirring of the larger particles will maintain a small residual velocity dispersion in the system (i.e. a pressure).

This dispersion induces a small but non-zero thickness of the disk, and is actually necessary to ensure the dynamical stability of the disk versus another destabilizing effect, namely self-gravity.

To quantify this effect, let us describe a planetary ring as a very thin disk where the surface density, velocity and pressure at a given point $\mathbf{r}$ and a given time t will be denoted respectively Σ , $\mathbf{v}$ and p . Furthermore, let Φ_D be the gravitational potential (per unit mass) created by the disk and c_s the speed of sound in the ring, corresponding to the typical velocity dispersion of the particles.

A detailed account of the calculations derived below and their physical interpretations can be found in [10]. In a first step, one can use a toy model where the disk is in *uniform* rotation with constant angular velocity vector $\mathbf{\Omega}$ directed along the vertical axis Oz . This will highlight the role of rotation in the stability of the ring, while simplifying the equations of motion. In reality, a disk around a planet exhibits a roughly Keplerian shear ($\Omega \propto r^{-3/2}$), but the conclusions concerning the role of rotation will not be altered in the present, simplified, approach.

The eulerian equations describing the dynamics of the disk are then, in a frame *rotating at angular velocity* $\mathbf{\Omega}$:

$$\begin{cases} \frac{\partial \mathbf{v}}{\partial t} + (\mathbf{v} \cdot \nabla) \mathbf{v} = -\frac{\nabla p}{\Sigma} - \nabla(\Phi_P + \Phi_D) - 2\mathbf{\Omega} \times \mathbf{v} + \Omega^2 \mathbf{r} \\ \frac{\partial \Sigma}{\partial t} + \nabla \cdot (\Sigma \mathbf{v}) = 0 \\ \nabla^2 \Phi_D = 4\pi G \Sigma \delta(z) \\ p = \Sigma c_s^2 \end{cases} \quad (1)$$

The first line is Euler's equation, where the acceleration on the fluid is expressed in terms of the pressure p , the potentials of the planet and the disk, Φ_P and Φ_D , respectively, the Coriolis term and the centrifugal acceleration. The second line the continuity equation, expressing the conservation of mass. The next line is Poisson's equation, relating the potential Φ_D of an infinitively thin disk to the surface density Σ , the gravitational constant G and the Dirac function $\delta(z)$, where z is the elevation perpendicular to the ring plane. Finally, the last line is the simplest equation of state, that of an isothermal disk, relating the pressure p to the surface density Σ through the speed of sound c_s .

A classical approach is to consider what happens to the disk when small perturbations are applied. For small enough disturbances, the equations above can be linearized, assuming that the unperturbed disk has uniform density. Then the various quantities involved can be decomposed into an unperturbed background value (with index 0) and a small perturbed value (with index 1): $\Phi = \Phi_P + \Phi_{D0} + \Phi_{D1}$, $\Sigma = \Sigma_0 + \Sigma_1$, $\mathbf{v} = 0 + \mathbf{v}_1$, $\Sigma = \Sigma_0 + \Sigma_1$.

This leads, after linearization of (1):

$$\begin{cases} \frac{\partial \Sigma_1}{\partial t} + \Sigma_0 (\nabla \cdot \mathbf{v}_1) = 0 \\ \frac{\partial \mathbf{v}_1}{\partial t} = -\frac{c_s^2}{\Sigma_0} \nabla \Sigma_1 - \nabla \Phi_{D1} - 2\mathbf{\Omega} \times \mathbf{v}_1 \\ \nabla^2 \Phi_{D1} = 4\pi G \Sigma_1 \delta(z) \end{cases} \quad (2)$$

One can seek how a given disturbance, for instance $\Sigma_1 = \bar{\Sigma}_1 \exp[i(\mathbf{k} \cdot \mathbf{r} - \omega t)]$, propagates in the disk, where $\bar{\Sigma}$ is the amplitude, $\mathbf{k}$ is the wavevector and ω is the frequency of the disturbance (with similar notations for the pressure and velocity). With no loss of generality we can assume that the disturbance propagates in the disk along the horizontal axis Ox , i.e. $\mathbf{k} = k\hat{\mathbf{x}}$.

Partial derivatives with respect to x and t are then equivalent to multiplications by ik and $-i\omega$, respectively. Also, Poisson's equation in the system 2 can be integrated as $|k|\bar{\Phi}_{D1} = -2\pi G\bar{\Sigma}_1$. This provides the algebraic system:

$$\begin{cases} ik\Sigma_0\bar{v}_{1x} & -i\omega\bar{\Sigma}_1 & = 0 \\ i\omega\bar{v}_{1x} & +2\Omega\bar{v}_{1y} + \left(\frac{2\pi G}{|k|} - \frac{c_s^2}{\Sigma_0}\right) ik\bar{\Sigma}_1 & = 0 \\ 2\Omega\bar{v}_{1x} & -i\omega\bar{v}_{1y} & = 0 \end{cases}$$

This system has non trivial solutions only if its determinant is non zero, thus providing the dispersion relation for propagating modes in a uniformly rotating disk:

$$\omega^2 = k^2 c_s^2 - 2\pi G \Sigma_0 |k| + 4\Omega^2 \quad (3)$$

Modes with $\omega^2 < 0$ will be unstable, so that the dispersion relation tells us which wavenumbers are unstable in the disk.

It is instructive to vary the values of c_s and Ω to see the respective effects of pressure and rotation on the stability of a thin (i.e. 2-D) disk. Various cases can be considered:

- *Cold and motionless disk*: $c_s = 0$ and $\Omega = 0$. Then $\omega^2 = -2\pi G \Sigma_0 |k|$, see Fig. 2(a), and all the modes are unstable, as expected, since nothing can stop the gravitational collapse of any disturbance. Note that the free fall collapse time, $t_{\text{colla}} \sim 1/\sqrt{G\Sigma_0|k|}$ depends on the wavenumber k of the disturbance, the smaller structures collapsing faster than the larger ones. This contrasts with the 3-D case, where the free fall time scale depends only upon the unperturbed density ρ_0 : $t_{\text{colla}} \sim 1/\sqrt{G\rho_0}$.
- *Hot and motionless disk*: $c_s \neq 0$ and $\Omega = 0$. Now $\omega^2 = k^2 c_s^2 - 2\pi G \Sigma_0 |k|$, so that the disk is unstable for disturbances with $|k| < k_J$, where

$$k_J = 2\pi G \Sigma_0 / c_s^2$$

is the Jeans wavenumber, see Fig. 2(b). The relation $|k| < k_J$ is actually the Jeans criterium for a 2-D disk. It can be compared to its 3-D classical counterpart $|k| < k_J = \sqrt{4\pi G \rho_0 / c_s^2}$. This criterium indicates how small the wavenumber must be (i.e. how *large and massive* the disturbance must be) in order to overcome the pressure associated with c_s .

- *Cold rotating disk*: $c_s = 0$ and $\Omega \neq 0$. This yields $\omega^2 = -2\pi G \Sigma_0 |k| + 4\Omega^2$, which shows that now the *large* disturbances (with small k) are stabilized by rotation, while the small structures ($|k| > k_R = 2\Omega^2 / \pi G \Sigma_0$) remain

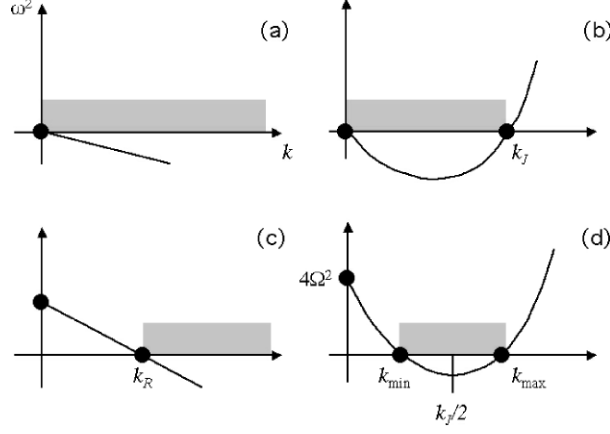


Fig. 2. Various cases of relation dispersions of free modes in rotating disks (3). In all panels, the grey intervals denote unstable mode regions. (a): motionless cold disk, (b): motionless hot disk, (c): rotating cold disk and (d): rotating hot disk. See text for details

unstable, see Fig. 2(c). This can be understood as large structures have a differential velocity in the disk due to the angular velocity Ω of the latter. As a consequence, a large structure which attempts to collapse under its own gravity will be stopped by the rotational barrier, when the centrifugal acceleration at its periphery balances its self gravity.

Note in passing that rotation does *not* stabilize the disk against all disturbances, as the smallest of them are still subject to gravitational collapses.

- *Hot rotating disk:* $c_s \neq 0$ and $\Omega \neq 0$. The dispersion relation $\omega^2 = k^2 c_s^2 - 2\pi G \Sigma_0 |k| + 4\Omega^2$ shows that ω^2 is minimum for $|k| = k_J/2 = \pi G \Omega_0 / c_s^2$, and that the minimum value is $\omega_{\min}^2 = 4\Omega^2 - \pi^2 G^2 \Sigma_0^2 / c_s^2$, as illustrated in Fig. 2(d).

It is convenient to define a dimensionless parameter, called the *Toomre parameter* ([9]):

$$Q = \frac{c_s \Omega}{\pi G \Sigma_0},$$

so that $\omega_{\min}^2 = (4\Omega^2 / Q^2) \cdot [Q^2 - 1/4]$. Thus, if $Q < 1/2$, then $\omega_{\min}^2 < 0$ and the disk remains unstable for an interval of wavenumbers $(k_{\min}, k_{\max})$ centered on $k_J/2$, see Fig. 2(d). Large structures (with small $|k|$) are then stabilized by rotation, while small structure (with large $|k|$) are stabilized by pressure. For

$$Q = \frac{c_s \Omega}{\pi G \Sigma_0} > \frac{1}{2},$$

the disk is stable to *any* disturbances. The above condition is called the Toomre criterium. Note that it is derived for a uniformly rotating disk. A

differentially rotating disk will also yield a similar criterium, except for the numerical coefficient in the right-hand side of the inequality.

Now, if a disk is rotating around a planet with potential Φ_P , then the equations of motion read (in a fixed frame this time):

$$\begin{cases} \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{v} = -\nabla (\Phi_P + \Phi_D) - \frac{\nabla \cdot p}{\Sigma} \\ \frac{\partial \Sigma}{\partial t} + \nabla(\Sigma \mathbf{v}) = 0 \\ \nabla^2 \Phi_D = 4\pi G \Sigma \delta(z) \\ p = \Sigma c_s^2 \end{cases} \quad (4)$$

It is then more convenient to work with the radial and tangential components of the velocity, v_r and v_θ , respectively, instead of v_x and v_y . The price to pay is that v_r and v_θ now depend on θ , and this must be accounted for when applying the nabla operator ∇ . The linearization of the (4) then proceeds as before, and the new dispersion relation is:

$$\omega^2 = k^2 c_s^2 - 2\pi G \Sigma_0 |k| + \kappa^2, \quad (5)$$

where $\kappa = r^{-3} \partial [r^3 \partial \Phi_P / \partial r] / \partial r$ is the so-called epicyclic frequency, basically the frequency at which a particle oscillates horizontally around its average position (note that κ coincides with the mean motion n in the special case of a Keplerian potential $\Phi_P \propto -1/r$).

Equation (5) shows that the disk is stable against any disturbances if:

$$Q = \frac{c_s \kappa}{\pi G \Sigma_0} > 1, \quad (6)$$

which constitutes a more general version of the Toomre criterium. It illustrates quantitatively how pressure (c_s) and rotation (κ) tend to stabilize the disk against self-gravity (Σ_0).

5 Particle Size and Ring Thickness

In planetary rings, inelastic collisions tend to reduce the velocity dispersion c_s . This in turn decreases the value of Q below unity, leading to gravitational instabilities at some point, according to (6), and see Fig. 3. This causes the ring to collapse into small lumps. At that point, the finite mass and size of the lumps will maintain a small but non-zero velocity dispersion, of the order of the escape velocity at the surface of the largest particles,

$$c_s \sim \sqrt{G m_{\max} / R_{\max}},$$

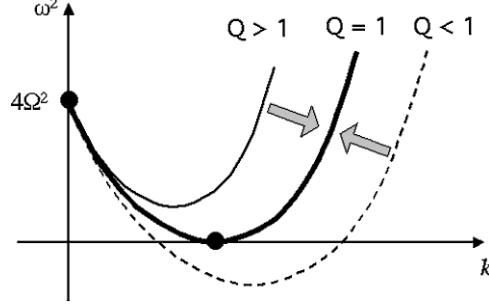


Fig. 3. Various cases of rotating hot disks: unstable ($Q < 1$), marginally stable ($Q = 1$) and stable for all modes ($Q > 1$). The arrows indicate that collisions and accretion tend to put the disk in the marginally stable state, see text

where $m_{\max}$ and $R_{\max}$ are the mass and radius of the largest lumps that dominate the dynamics of the ring. This allows the latter to maintain a Q value just above unity.

Conversely, if the lumps become too large, then the velocity dispersion increases and $Q > 1$, leading to an increase of dissipation and also to the disruption of the lumps. This will eventually decrease again the value of Q back to unity.

Thus, an equilibrium is reached, where a marginal stability is maintained ($Q \sim 1$), as illustrated in Fig. 3. This, combined with (6) and (5), leads to:

$$R_{\max} > \sim \left(\frac{a}{R_P} \right)^{3/2} \left(\frac{\Sigma_0}{100 \text{ g cm}^{-2}} \right) \left(\frac{\rho}{\text{g cm}^{-3}} \right)^{1/2} \text{ meters ,}$$

where a is the distance to the planet center and ρ is the bulk density of the largest particles.

As an example, if we take typical values of $\Sigma_0 \sim 20 \text{ g cm}^{-2}$ corresponding to Saturn's A ring, $\rho < \sim 1 \text{ g cm}^{-3}$ for icy particles and $a \sim 2R_P$, this yields $R_{\max} > \sim$ of a few meters, in accordance with radio *Voyager* observations.

More generally, the thickness h of the ring is given by $h \sim c_s/\Omega$, where the orbital angular velocity $\Omega \sim \kappa \sim \sqrt{GM_P/a^3}$ in a Keplerian disk. Also, a homogeneous ring of radius a has a mass of $m_r \sim \pi \Sigma_0 a^2$. Combining these relations with the expression of Q yields:

$$\frac{h}{a} \sim Q \times \frac{m_r}{M_P} . \quad (7)$$

Thus, we see that the thickness of a marginally stable ring is eventually imposed by the ratio of the ring mass to the planet mass: the extreme thinness of planetary rings comes from their extremely small mass when compared to the planet. For Saturn's A ring, with $m_r/M_P \sim 10^{-7}$, we obtain $h \sim$ a few meters, in agreement again with (indirect) measurements of h , see Sect. 7.

Note for closing this section that the first marginally unstable modes, appearing when $Q \sim 1$, corresponds to the minimal value of ω^2 in (5). They have wavenumbers $k_{\text{unst}} \sim k_J/2 \sim \Omega/c_s$, i.e. wavelengths:

$$\lambda_{\text{unst}} \sim 2\pi h ,$$

or a few tens of meters. These marginal instabilities are probably the explanation for the quadrant asymmetries observed in Saturn's A ring.

6 Resonances in Planetary Rings

So far, we have been considering *free* modes propagating in rings. We now turn to the case where modes are *forced* by a satellite near a resonance.

Considering the very small masses of the satellites relative to the giant planets, these forced modes are in general “microscopic”, in the sense that they induce deviations of a few meters on the particles orbits. However, near resonances, a satellite can excite macroscopic responses in the disk, which exhibit large collective disturbances over tens of kilometers, i.e. on scales observable by spacecraft imagers.

As explained later, this collective response allows an secular exchange of angular momentum and energy between the ring and the satellite, very much like tides allow secular exchanges between a planet and its satellites.

The equations of motion are the same as in (4), except that the first equation (Euler's) must now accounts for the forcing of the satellite through its disturbing potential Φ_S :

$$\left\{ \begin{array}{l} \left(\frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla \right) \mathbf{v} = -\nabla (\Phi_P + \Phi_D + \Phi_S) - \frac{\nabla \cdot \mathbf{P}}{\Sigma} \\ \frac{\partial \Sigma}{\partial t} + \nabla(\Sigma \mathbf{v}) = 0 \\ \nabla^2 \Phi_D = 4\pi \Sigma G \delta(z) \\ p = \Sigma c_s^2 , \end{array} \right. \quad (8)$$

Note in passing that we have replaced the (scalar) isotropic pressure p by a pressure tensor, $\mathbf{P}$. This allows us to take into account in a general way more complicated effects like viscosity or non-isotropic pressure terms.

The simplest case we can think of is the forcing of a homogeneous disk by a small satellite of mass $m_s \ll M_P$ with a circular orbit of radius a_s and mean motion n_s around a point-like planet with potential $\Phi_P(r) = -GM_P/r$. This is the Keplerian approximation. In reality, the oblateness of the planet introduces extra terms causing a slow precession of the apse and node of the orbit. These subtleties will not be taken into account here since they obscure our

main point (the study of a simple isolated resonance) without changing our main conclusions. These effects would become important, however, if the orbital eccentricity and inclination of the satellite were to come into play.

With these assumptions, the satellite potential is periodic in $\theta - n_S t$, so that it can be Fourier expanded as:

$$\Phi_S(r, \theta, t) = \sum_{m=-\infty}^{+\infty} \Phi_{Sm}(r) \cdot \exp[im(\theta - n_S t)] , \quad (9)$$

where m is an integer, $\Phi_{Sm}(r) = -(Gm_S/2a_S) \cdot b_{1/2}^m(r/a_S)$ and $b_{1/2}^m$ is the classical Laplace coefficient (see [12] for a review of the properties of these coefficients). We note that $\Phi_m(r) = \Phi_{-m}(r)$, since Φ_s is real.

Equations (8) are then linearized, and we assume that the free modes of the disk are damped by collisions, or are at least negligible with respect to the forced modes, especially near the resonances. Then all the perturbed quantities, for instance the radial velocity v_r , take the same form as the forcing (9), i.e. $v_r(r, \theta, t) = \sum v_{rm}(r) \cdot \exp[jm(\theta - n_S t)]$, etc. . .

Each term $\Phi_{Sm}(r) \cdot \exp[jm(\theta - n_S t)]$ of the satellite potential then forces a mode in the ring, and if equations (8) remain linear, then it is enough to study the reaction of the disk to each mode separately. As already noted before, this replaces the differential operators by mere multiplications, namely $\partial/\partial t = -jmn_s$ and $\partial/\partial \theta = jm$.

This approach is especially useful near resonances, where one mode dominates over all the other ones, and can thus be “clipped off” from the rest. Consider a particle with mean motion n , so that its longitude writes $\theta = nt$ (plus an arbitrary constant). This particle thus feels a forcing potential $\Phi_{Sm}(r) \cdot \exp[jm(\theta - n_S t)] = \Phi_{Sm} \exp[jm(n - n_S)t]$, i.e. a term with frequency $m(n - n_S)$.

A so-called (“Excentric Lindblad Resonance”) (ELR) occurs when this frequency matches the horizontal epicyclic frequency κ of the particle, i.e. when:

$$\kappa = \pm m(n - n_S) . \quad (10)$$

In this case, and for small horizontal displacements of the particle, the latter behaves very much like a harmonic oscillator (i.e. a linear system) near a classical resonance. This simple model predicts that the horizontal displacement of the particle increases as it approaches the resonance, and becomes singular at exact resonance¹. In reality, non-linear terms come into play in the disk and eventually prevent the singularity. In counterpart, this complicates significantly the equations of motion, and renders the system rather untractable.

Fortunately, for sufficiently dense planetary rings perturbed by very small satellites, collisions, pressure and self-gravity prevent a wild behavior for

¹This horizontal displacement is proportional to the orbital eccentricity of the particle, hence the nomenclature “Excentric Lindblad Resonance”

nearby streamlines, thus keeping the perturbed motion small, and eventually ensuring that the system 8 remains linear.

Note that for a Keplerian disk $\kappa = n$, so that the condition $\kappa = \pm m(n - n_S)$ is equivalent to

$$n = \frac{m}{m \mp 1} \cdot n_S , \quad (11)$$

corresponding to mean motion resonances 2:1, 3:2, 4:3, etc..., also called first order resonances. Other resonances, e.g. second order resonances 3:1, 4:2, 5:3 (i.e. of the form $m : m - 2$) can also come into play when the smaller second order terms in the particle orbital eccentricity are considered. Still other resonances (referred to as “corotation resonances”) can also arise when the satellite orbital eccentricity is accounted for. These kinds of resonances fall outside the main topic of this chapter and will not be considered here.

Another important simplification comes from the fact that in planetary rings, the perturbed quantities vary much more rapidly radially than azimuthally. Physically, this means that the spiral structures resonantly forced are *tightly wound*, like the grooves of a music disk. More precisely, the lower order radial derivatives can be neglected with respect to higher orders:

$$\frac{m^2}{r^2} \ll \frac{m}{r} \cdot \frac{d}{dr} \ll \frac{d^2}{dr^2} . \quad (12)$$

This is the WKB approximation², which greatly simplifies the system 8, leading to (see [13]):

$$\left\{ \begin{array}{lcl} jm(n - n_s)v_{rm} - 2nv_{\theta m} & = & -\dot{\Phi}_{Sm} - \dot{\Phi}_{Dm} - \frac{c_s^2 \dot{\sigma}_m}{\Sigma_0} + \left(\mu + \frac{4\nu}{3} \right) \ddot{v}_{rm} \\ \frac{n}{2}v_{rm} + jm(n - n_s)v_{\theta m} & = & -jm \frac{\Phi_{Sm} + \Phi_{Dm}}{r} - jm \frac{c_s^2 \sigma_m}{r \Sigma_0} + \nu \ddot{v}_{\theta m} \\ \sigma_m & = & -\frac{\Sigma_0 \dot{v}_{rm}}{jm(n - n_s)} \\ \dot{\Phi}_{Dm} & = & -2\pi G j s \sigma_m \\ p_m & = & c_s^2 \sigma_m . \end{array} \right. \quad (13)$$

The quantities μ and ν are the bulk and shear kinematic viscosities, respectively, coming from the pressure tensor $\mathbf{P}$. The dot stands for the *space* (not time) derivative d/dr . The Poisson equation has been solved using the results of [12], where $s = \pm 1$ is chosen in such a way that the disk potential out of the disk plane tends to zero:

$$\Phi_{Dm}(r + is|z|) \rightarrow 0 , \quad (14)$$

²Developed by Wentzel-Kramer-Brillouin in the field of quantum mechanics.

as $|z|$ goes to infinity. We will see that boundary conditions actually impose $s = +1$.

If we forget for the moment the terms Φ_{Dm} , σ_m , μ and ν , i.e. if we consider a test disk with no self-gravity, no pressure nor viscosity, then we get:

$$\begin{cases} v_{rm}(r) = -jm \left[(n - n_s)r \frac{d}{dr} + 2n \right] \cdot \Phi_{Sm}(r) \frac{1}{rD} \\ v_{\theta m}(r) = \left[nr \frac{d}{dr} + 2m^2(n - n_s) \right] \cdot \Phi_{Sm}(r) \frac{1}{2rD} , \end{cases} \quad (15)$$

where $D(r) = n^2 - m^2(n - n_s)^2$ is a measure of the distance to exact resonance. The velocity is singular when $D = 0$, i.e. when $n = m/(m \mp 1)n_s$, corresponding to the condition (11). Thus the dependence in $1/D$ is just the expected response of a linear oscillator near a resonance.

The result obtained above does not strike by its simplicity: complicated equations and tedious calculations have just shown that a harmonic oscillator behaves as derived in basic text books. However, we have gained with these equations some important insights into more subtle effects associated with viscosity, pressure and self-gravity. More generally, these equations show how *collective* effects modify the simple harmonic oscillator paradigm into more complicated behaviors.

Near the resonance, $D = 0$, the system 13 is almost degenerate, and (15) yield $u_{\theta m} \sim \pm(j/2)u_{rm}$. To solve for u_{rm} , one uses this degeneracy, plus the tightly wound wave condition (12). We note x the relative distance to the resonance radius a_m , $x = (a - a_m)/a_m$, and we expand (13) near $x = 0$, which yields:

$$-\alpha_v^3 \frac{d^2}{dx^2}(u_{rm}) + \alpha_G^2 \frac{d}{dx}(u_{rm}) - jxu_{rm} = C_m , \quad (16)$$

where:

$$\begin{cases} \alpha_v^3 = j\alpha_P^3 + \alpha_\nu^3 \\ \alpha_P^3 = \mp \frac{c_s^2/n}{3ma_m^2 n_s} \\ \alpha_\nu^3 = \frac{\mu + 7\nu/3}{3ma_m^2 n_s} \\ \alpha_G^2 = \pm \frac{2\pi s G \Sigma_0}{3ma_m n n_s} , \end{cases} \quad (17)$$

and C_m is a factor which weakly depends on m ([13]). For purposes of numerical applications, $C_m \sim \pm 0.27an(m_s/M)$ as m tends to infinity. The coefficient $C_m \propto m_s$ in (16) is the forcing term due to the satellite. The coefficients α_P , α_ν and α_G encapsulate the effects of pressure, viscosity and self-gravitation, respectively. In the absence of all the α 's, the response of the disk is indeed

singular at the resonance $x=0$: $u_{rm} \propto 1/x$, as expected in a test disk in the linear regime. The extra terms with the α 's in (16) prevent such an outcome, and forces the solution to remain finite at $x = 0$. If oscillations are present in the solution, then waves are launched.

Equation (16) can be solved by defining the Fourier transform of u_{rm} :

$$\tilde{u}_{rm}(k) = \int_{-\infty}^{+\infty} \exp(-jkx) u_{rm}(x) dx ,$$

assuming that $u_{rm}(x)$ is square integrable. Then we take the Fourier transform of (16):

$$\frac{d}{dk}(\tilde{u}_{rm}) + (\alpha_v^3 k^2 + j\alpha_G^2 k) \tilde{u}_{rm} = 2\pi C_m \delta(k) , \quad (18)$$

where δ is the Dirac function. This first-order equation is solved with the boundary condition $\tilde{u}_{rm} \rightarrow 0$ as $k \rightarrow \infty$, since $\tilde{u}_{rm}$ is a Fourier transform. Then:

$$\tilde{u}_{rm}(k) = 2\pi C_m H(k) \exp[-(\alpha_v^3 k^3/3 + j\alpha_G^2 k^2/2)] ,$$

where H is the unit-step function ($=0$ for $k < 0$ and $=1$ for $k > 0$). This eventually provides the solution we are looking for:

$$u_{rm}(x) = C_m \int_0^{+\infty} \exp[j(kx - \alpha_G^2 k^2/2 - \alpha_P^3 k^3/3) - \alpha_v^3 k^3/3] dk . \quad (19)$$

Note that the boundary condition (14) also requires $u_{rm}(x + is|z|) \rightarrow 0$ as $|z| \rightarrow +\infty$, i.e. $s = +1$ since $k > 0$ in the integral above.

The qualitative behavior of $u_{rm}(x)$ can be estimated from the behavior of the argument in the exponential, $j(kx - \alpha_G^2 k^2/2 - \alpha_P^3 k^3/3) - \alpha_v^3 k^3/3$. This argument has an imaginary part, $j(kx - \alpha_G^2 k^2/2 - \alpha_P^3 k^3/3)$, which causes an oscillation of the function in the integral, and a real part, $-\alpha_v^3 k^3/3$, which causes a damping of that function.

The integral in (19) is significant only when the phase $kx - \alpha_G^2 k^2/2 - \alpha_P^3 k^3/3$ is stationary, i.e. near the wave number k_{stat} such that:

$$\alpha_G^2 k_{\text{stat}} + \alpha_P^3 k_{\text{stat}}^2 = x \quad (20)$$

somewhere in the domain of integration ($k > 0$).

For instance, if the disk is dominated by self-gravity, i.e. $\alpha_G^2 \gg \alpha_P^3$, then the condition (20) reduces to $x = \alpha_G^2 k_{\text{stat}}$. Thus, the integral in (19) is significant only on that side of the resonance where x and α_G^2 have the same sign. In that case, the solution of 19 oscillates near x with a local radial wave number $k_{\text{stat}} \approx x/\alpha_G^2$. The local radial wavelength of the wave is thus $\propto 1/x$. Consequently, the wave oscillates more and more and more rapidly as it propagates away from the resonance, see Fig. 4(a). On the other side of the resonance (where x and α_G^2 have opposite signs), the argument of the exponential in (19) is never stationary, and the integral damps to zero. This means that the

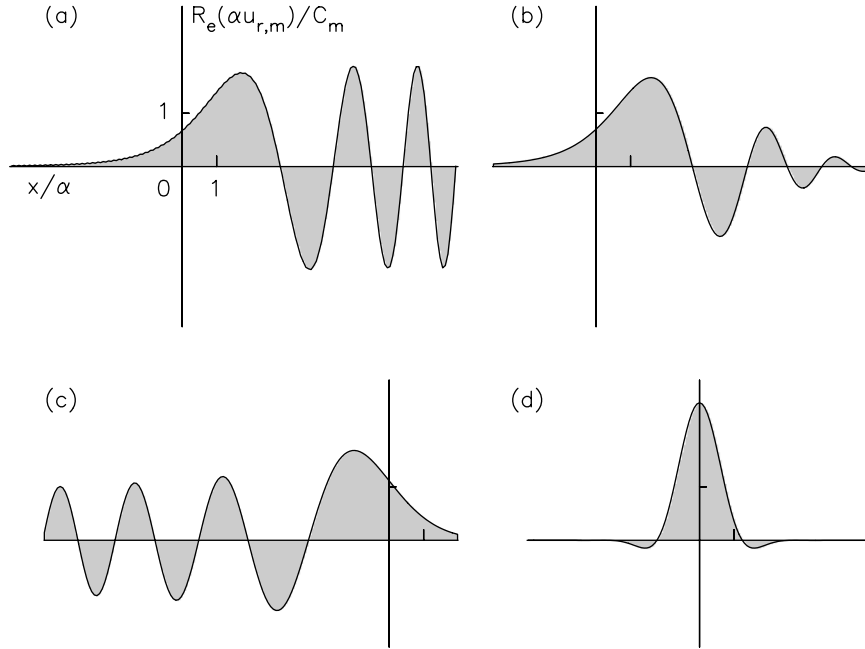


Fig. 4. Various responses of a disk near an inner Lindblad resonance (located at $x = 0$). The term α which appears in the definition of the abscissa and ordinate units represents any of the coefficients defined in (17), depending on the case considered. (a) A disk dominated by self-gravity. The wave is launched at $x = 0$ and propagates to the right of the resonance, while remaining evanescent on the left side. (b) A self-gravity wave damped by viscosity. (c) A wave in a disk dominated by pressure. The propagating and evanescent sides are inverted with respect to the self-gravity case. (d) Response in a disk dominated by viscosity. The wave is now evanescent on both sides of the resonance

wave is evanescent, with a typical damping distance of $\sim |\alpha_G|$ in the forbidden region, see again Fig. 4(a).

The same reasoning shows that when the disk is dominated by pressure ($\alpha_P^3 \gg \alpha_G^2$), then the wave propagates on the side of the resonance where x and α_P^3 have the same sign. The local radial wave number is now $k_{\text{stat}} \approx \sqrt{x/\alpha_P^3}$, and the local radial wavelength is $\propto 1/\sqrt{x}$. Again the wave oscillates more and more rapidly as it goes away from the resonance, but not so drastically as for a wave supported by self-gravity, see Fig. 4(c). When x and α_P^3 have opposite signs, the wave is evanescent over a damping distance of $\sim |\alpha_P|$.

The effect of viscosity is illustrated in Figs. 4(b) and 4(d). In case (b), viscosity remains weak enough to allow for a few self-gravity waves to propagate in the disk. In case (d), viscosity completely dominates the disk response, and no wave can be launched from the resonance.

7 Waves as Probes of the Rings

It is interesting to compare the numerical values of the coefficients α_G and α_P for planetary rings. The larger of the two coefficients tells us which process (self-gravity or pressure) dominates in the wave propagation. According to the expressions given in (17), using the definition of Toomre's parameter, $Q = c_s n / \pi G \Sigma_0$, and remembering that the thickness of the ring is given by $h \sim c_s / n$, we obtain:

$$\left| \frac{\alpha_P}{\alpha_G} \right| \sim \sqrt{2Q} \left(3m \frac{h}{a} \right)^{1/6}.$$

As we saw before, $h/a \sim 10^{-7}$ is very small and $Q \sim 1$, while m is typically a few times unity. Thus, the ratio $\alpha_P/\alpha_G \sim 0.1 - 0.2$ is small, but not by an overwhelming margin, because of the exponent $1/6$ in the expression above.

The same is true with the ratio α_ν/α_G since $\alpha_\nu \sim \alpha_P$. This is because the kinematic viscosities μ and ν are both of the order of c_s^2/n is moderately thick planetary rings ([2]).

Consequently, self-gravity is the dominant process governing the propagation of density waves in planetary rings, but viscosity is efficient enough to damp the wave after a few wavelengths, see for instance the panel (b) in Fig. 4.

Note that self-gravity waves are macroscopic features which can be used as a probe to determine microscopic parameters such as the local surface density Σ_0 of the ring, or its kinematic viscosities μ or ν . This method has been used with bending waves in Saturn's A ring and is the only way so far to derive Σ_0 or ν in these regions ([5]).

The determination of ν has an important consequence, namely the estimation of the local thickness h of the ring, since $\nu \sim c_s^2/n$. Typical values obtained for Saturn's A ring indicate that $h \sim 10 - 50$ meters, a value already consistent with stability considerations, see for instance the discussion after Eq. (7).

8 Torque at Resonances

A remarkable property of the function $u_{rm}(x)$ defined in (19) is that the real part of its integral, $\Re[\int_{-\infty}^{+\infty} u_{rm}(x)dx]$, is *independent* of the values of the coefficients α 's. For instance, all the areas under the curves of Fig. 4 (i.e. the shaded regions) are equal, *including* in the cases (c) and (d), where dissipation plays an important role.

This can be shown by using an integral representation of the step function ([13]),

$$H(k) = (j/2\pi) \int_{-\infty}^{+\infty} [\exp(-jku)/(u + j\epsilon)]$$

in (19), where ϵ is an arbitrarily small number. Equation (19) may then be then integrated in x , which yields $\delta(k)$, then in k , and finally in u :

$$\Re \left(\int_{-\infty}^{+\infty} u_{rm} dx \right) = \Re \left(j C_m \int_{-\infty}^{+\infty} \frac{du}{u + j\epsilon} \right) = \pi C_m . \quad (21)$$

Now, the complex number $u_{rm}(x)$ describes how the disk responds to the resonant excitation of the satellite at the distance x from the resonance. More precisely, the modulus $|u_{rm}(x)|$ is a measure of the amplitude of the perturbation at x , and is thus directly proportional to the eccentricity of the streamlines around x . The argument $\phi = \arg[u_{rm}(x)]$ is on the other hand directly connected to the phase lag Ψ of the perturbation with respect to the satellite potential. It can be shown easily that $\phi = \Psi \mp \pi/2$, see [13].

Consequently, the satellite torque acting on a given streamline is proportional to its eccentricity $\propto |u_{rm}(x)|$ and to $\sin(\Psi) \pm \cos(\phi)$, a classical properties of linear oscillators. Consequently the total torque exerted at the resonance is proportional to $\Re[\int_{-\infty}^{+\infty} u_{rm}(x) dx]$.

More precisely, the torque exerted by the satellite on the disk is by definition $\Gamma = \int \int (\mathbf{r} \times \nabla \Phi_S) \Sigma d^2r$, where Φ_S and Σ may be Fourier expanded according to (9) when the stationary regime is reached. After linearization, one gets the torque exerted at the resonance:

$$\Gamma_m = \mp 12\pi m^2 \Sigma_0 a_s^2 C_m \Re \left(\int_{-\infty}^{+\infty} u_{rm}(x) dx \right), \quad (22)$$

where the upper (resp. lower) sign applies to a resonance inside (resp. outside) the satellite orbit.

This is the so-called “standard torque” ([1]), originally derived for a self-gravity wave launched at an isolated resonance. The calculations made above show that this torque is actually *independent* of the physical process at work in the disk, as long as the response of the latter remains linear. In particular, dissipative processes such as viscous friction do not modify the torque value.

This torque allows a secular exchange of angular momentum between the disk and the satellite. Note that the sign of this exchange is such that the torque always tends to push the satellite away from the disk.

This torque have a wide range of applications that we will not review here. We will just note here that it may lead to the confinement of a ring when two satellites lie on each side of the latter (the so-called “shepherding mechanism”). This could explain for instance the confinement of some of the narrow Uranus’ rings.

Another consequence of such a torque is that Saturn’s rings and the inner satellites are continuously pushed away from each other. The time scales associated with such interactions (of the order of 10^8 years) tend to be shorter than the age of the solar system ([2]). This suggests that planetary rings are either rather young, or, if primordial, have continuously evolved and lived several cycles since their formation.

9 Concluding Remarks

We have considered in this chapter some fundamental concepts associated with rings: their flattening, their thickness and their resonant interactions with satellites. Note that these processes are mainly linked to the larger particles of the rings. Furthermore, they make a useful bridge between the microscopic and macroscopic properties of circumplanetary disks.

Meanwhile, many other processes have not been discussed here, such as the effect of electromagnetic forces on dust particles, the detailed nature of collisions between the larger particles, the accretion and tidal disruption of loose aggregates of particles, the origin of sharp edges in some rings, their normal modes of oscillation, etc. . .

These issues, and others, are addressed in some of the references given in the bibliography below. All the processes involved clearly show that rings are by no means the simple and everlasting objects they seem to be when observed from far away.

References

1. P. Goldreich, S. Tremaine: *Ann. Rev. Astron. Astrophys.* **20**, 249 (1982)
2. N. Borderies, P. Goldreich, S. Tremaine: unsolved problems in planetary ring dynamics. In: *Planetary rings*, ed by R. Greenberg, A. Brahic (Univ. Of Arizona Press 1984) pp 713–734
3. P.D. Nicholson, L. Dones: *Rev. Geophys.* **29**, 313 (1991)
4. P. Goldreich: puzzles and prospects in planetary ring dynamics. In: *Chaos, resonance and collective dynamical phenomena in the solar system*, ed by S. Ferraz-Mello (Kluwer Academic Publishers, Dordrecht Boston London 1992) pp 65–73
5. L.W. Esposito: *Annu. Rev. Earth Planet. Sci.* **21**, 487 (1993)
6. J.N. Cuzzi: *Earth, Moon, and Planets* **67**, 179 (1995)
7. A.W. Harris: the origin and evolution of planetary rings. In: *Planetary rings*, ed by R. Greenberg, A. Brahic (Univ. Of Arizona Press 1984) pp 641–659
8. W. Ward: the solar nebula and the planetesimal disk. In: *Planetary rings*, ed by R. Greenberg, A. Brahic (Univ. Of Arizona Press 1984) pp 660–684
9. A. Toomre: *ApJ.* **139**, 1217 (1964)
10. J. Binney, S. Tremaine: *Galactic Dynamics* (Princeton University Press 1988)
11. P.Y. Longaretti: Planetary ring dynamics: from Boltzmann’s equation to celestial dynamics. In: *Interrelations between physics and dynamics for minor bodies in the solar system*, ed by D. Benest, C. Froeschlé (Editions Frontières, Gif-sur-Yvette 1992) pp 453–586
12. F.H. Shu: waves in planetary rings. In: *Planetary rings*, ed by R. Greenberg, A. Brahic (Univ. of Arizona Press 1984) pp 513–561
13. N. Meyer-Vernet, B. Sicardy: *Icarus* **69**, 157 (1987)
14. B. Sicardy: Planetary ring dynamics: Secular exchange of angular momentum and energy with a satellite. In: *Interrelations between physics and dynamics for minor bodies in the solar system*, ed by D. Benest, C. Froeschlé (Editions Frontières, Gif-sur-Yvette 1992) pp 631–651

Index

- A. Clairaut 96
- Abel-Jacobi 16
- accelerometers 83
- accretion 118
- Al Kashi 72
- Andoyer's
 - modified variables 161
 - variables 160
- ANEOs 125
- angular momentum 6, 8, 9, 186
 - conservation of 91
- angular velocity 11
- artificial
 - satellites 68, 70, 76
- asteroid 88, 89, 92, 117, 118, 136
 - 1999 AN₁₀ 146
 - Ausonia 101
 - belt 138, 139
 - binary 98, 100, 103
 - C-type 100
 - comet P/Halley 91
 - damping 92
 - Eunomia 111
 - family 133, 136
 - Hektor 111
 - Ida 103
 - Kalliope 105
 - Kleopatra 111
 - lightcurve 89, 98
 - main belt 136
 - porosity 98
 - rotation 90
 - rubble-pile 95, 98
 - S-type 99
 - satellites 135
 - Toutatis 148, 158
- atmospheric loading 37
- Ausonia 109
 - asteroid 100
- axes
 - principal 4, 5
- axis of rotation 60
- b*-plane 145, 147, 153
 - crossing 154
 - post-encounter 155
- BepiColombo 170
- binary systems 103
- body
 - celestial 12
 - parent 133
 - rigid 7
 - spinning 1
 - tide 36
- Bruno Sicardy 182
- bulk density 103, 105
- Cassini 159
 - division 184
 - state 159, 163
- Ceres 95
- CHAMP 82
 - mission 83
- Chandler 30
- Chandrasekhar 96
- Clairaut 104
- close encounters 145

- planetary 146
- CMB (core mantle boundary) 38
- CMB topography 45
- collisions 117, 185
 - catastrophic 118
 - disruptive 121
 - inelastic 190
 - low-velocity 118
- comet P/Halley 91
- conservation laws 15
- conservation of energy 25
- conservation of mass 29
- conservation of momentum 29
- convection 27, 28, 46
- coordinates 153
 - angles-actions 167
 - post-encounter 155
- Coriolis 187
- crack growth velocity 127
- D'Alembert 49
- D. Hestroffer 88
- damping timescale 94, 95
- Darwin 30
- deformable stratified planet 49
- deformation 23, 28
- density 70
 - bulk 99, 101
- diffusion 25, 137
- Dirac
 - function 187
- disk
 - cold motionless 188
 - cold rotating 188
 - dynamics of the 187
 - flat 186
 - hot motionless 188
 - hot rotating 189
 - Keplerian 194
 - potential 194
 - stability 186
- dispersion
 - of fragments 135
 - relation 190
 - velocity 191
- disruption 120, 121, 123, 126, 136
 - asteroid 128
 - catastrophic 133
 - of monolithic family parent bodies 134
 - tidal 184
- dissipation 59, 92
 - internal 159
- dissipative forces 165
- distribution
 - fragment size 120, 122
 - fragment velocity 122
 - size-velocity 123
- DORIS 79
- dynamical flattening 75
- dynamics of rotation 6
- Earth 74
 - centre of figure 76
 - gravity field 67, 77, 79, 81
 - "mean" radius 70
 - non-rigid spheroidal 102
 - rotating 95
 - shape of the 71
 - structure of the 68
 - tensor of inertia of the 76
 - variations in rotation 85
- eigen values 179
- ejection velocity 122, 139
- El Niño 81
- elastic 9
- elastic deformation 38, 46
- elastic energy 44
- elastic-plastic
 - body 102
 - theories 101
- elasto-gravitational deformation 31
- elasto-gravitational equation 33
- electromagnetic coupling 57
- ellipsoids
 - tri-axial 96, 98, 102, 103
- Encke
 - division 184
- encounter 154
- energy 6
 - conservation of 17
 - potential 17
 - rotational 6
- entropy 25
- ENVISAT 82
- equation of rotation 51
- equations

- conservation 124
- Euler's 11, 14, 16
- of dynamics 70
- Poisson's 187
- variational 80
- equations of state 26
- equilibrium
 - elastostatic 101
 - figures 95
 - hydrostatic 96, 97, 103
 - static 102
- equipotential surfaces 70
- ERS 82
- Eugenia 100
 - asteroid 99
- Euler 49
 - angles 90–92, 160
 - equations 12, 14, 90, 91
 - frame 130
 - top 13–15
- Eunomia 138
 - family 134
- excitation function 58
- families
 - asteroid 119
- FCN (free core nutation) 58
- FGS 107
 - data 111
 - interferometer 105
 - measurements 111
 - sensitivity 107
- FICN (free inner core nutation) 58
- field 67
- figure of equilibrium 95
- Fine Guidance Sensor (FGS) 106
- flattening of the Earth 54
- Florent Deleflie 67
- fluid
 - core 50, 75
 - homogeneous 97
 - non homogeneous 101
 - uniformly rotating 97
- force
 - centrifugal 10
 - non gravitational 83
- Fourier transform 39, 196
- fragmentation
 - degree of 121
 - phase 119, 130
- fragments 138
- free modes 192
- friction 19, 103, 105
 - coefficient 11
- functions
 - associated 73
 - elliptic 16, 17
 - holomorphic 16
 - Jacobian elliptic 15
 - theta 16
- Gauss'
 - formulae 135
- generalized setup 149
- geocentre
 - motion 77, 85
 - position 77
- geodesy 69, 70
 - space 70
- geoid 71, 85
- GEOSAT 82
- Giovanni B. Valsecchi 144
- glaciation/deglaciation cycle 45
- GOCE 83–85
- GPS 79, 83, 85
- GRACE 83, 85
- Grady-Kipp 129
 - fracture model 131
 - fragmentation theory 126
- gravitational phase 131, 132
- gravitational potential 34, 43, 59
- gravitational torque 55
- gravity 17, 19, 71
 - potential of 71
- gravity field 67, 68, 71, 72, 75, 76, 78, 81, 84, 85
 - coefficients of 79
 - global 75, 82
 - static 76
- Greff M. 23–66
- GRIM5-C1 82
- GRIM5-S1 model 82
- gyroscopes 11, 12
- Hamiltonian 162, 165, 173
 - formalism 171
 - function 165
 - of the free body rotation 161

- quadratic 166
- simplified 177
- Hebe
 - asteroid 90
- Helmotz equation 53
- Helmotz vortex 53
- herpolhode 13
- historical review 23
- homogeneous incompressible planet 43
- Hooke
 - law 39, 101, 102, 125
- Hopkins 30, 50
- HST
 - displacement 107
 - observations 109
 - “orbit” 107
- HST/FGS 89, 105, 111, 113
- Hubble Space Telescope (HST) 105
- Hugoniot 125
 - elastic limit 126
- hydrocode 119, 124
 - Lagrangian 130
- hydrostatic equilibrium 46, 55, 96, 100, 103
- hydrostatic figure of the Earth 27
- ICB (inner core boundary) 53
- impact
 - cratering 121
 - energy 121, 123, 140
 - velocity 118
- inertia 3
 - matrix 3, 4, 11, 19
 - tensor 36, 52, 96
- IRAS 105
- Jacobi 96
 - ellipsoid 98–101
 - figures 99
 - integrals 97
 - sequence 97, 98
- Jacques Henrard 159
- Jameson 30
- Jeans
 - criterium 188
- JPL 83
- Karin 137
- family 137
- Kelvin 30
- Kelvin body 41
- Kepler’s
 - third law 103
- Keplerian motion 186
- keyholes 146
- kinetic energy 3
 - conservation of 91
- Koester
 - prism 106
- Koronis
 - family 134, 137, 138
- Kovalevskaya’s top 20
- Laboratory Experiments 120
- LAGEOS-1 76, 81
- Lagrange equations 77, 104
- Lagrange’s top 17
- Lagrangian 32, 124
 - attraction 34
 - codes 130
 - frame 130
- Lamé
 - coefficients 101
- Laplace 100
- Laplace equation 69, 72
- laws
 - conservation 15, 17, 18
- Legendre polynomials 72
- Legros H. 23–66
- libration 159
 - forced 165
 - free 165
 - frequencies of 165
- lightcurve 89, 95, 100, 109
- linear rheology 39
- Liouville equation 51
- loading
 - glacial 75
- Love number 35
- Lunar Laser Ranging 76
- Mac Cullagh’s formula 36
- Maclaurin 96, 97
 - oblate spheroid 100
 - spheroid 101
- magnetic friction 59
- magnetic pressure 38

- mantle 75
- Mariner 10 169
- Mathilde
 - asteroid 103
- Maxwell body 41
- mean equatorial radius 74
- Mercury 169
 - rotation of 168
- MESSENGER 170
- meteorite 7
- Milankovitch 50
- Mohr-Coulomb
 - criteria 103
 - model 103
- MOID 148, 153
 - local 151, 152
- Molodensky 50
- moment of inertia 4, 5, 9, 15, 52
 - principal 4
- momentum 3, 6
 - angular 3
- moon
 - rotation of 159
- motion
 - in rotational 2, 3
 - in translational 2, 3
- MUSES-C
 - japanese mission 131
- N-body
 - codes 132
 - interactions 132
- Newcomb 30
- Newton 49, 95
 - law of gravitation 67
 - universal law of gravitation 68
- non elasticity
 - moon 165
- numerical integration 78
- numerical simulations 118, 119, 121, 133, 137
- nutation 18, 19, 92
- ONERA 83
- Öpik's theory 144, 148, 149
- P. Tanga 88
- Pallas 95
- Pan 184
- paragraph 183
- Patrick Michel 117
- phase
 - fragmentation 134
 - gravitational 134
- Phobos 103
- Pierre Exertier 67
- planetary convection 24
- planetary rings
 - dynamics of 182
- planetesimals 118
- planets interior 28
- Poincaré 50
- poinsot 13
 - picture 16
- Poisson
 - equation 31, 194
 - ratio 101
- polhode 13, 14
- porosity 95, 105
 - macro 99, 100
- post glacial rebound 76, 81
- post-encounter 156
 - b*-plane 155
- potential 69–72, 161
 - development of the gravitational 174
 - gravity 68
 - Keplerian 190
 - satellite 193
- precession 15, 19, 92
 - fast 19
 - periastron 104
 - pure 18
 - slow 19
- PREM (model) 24
- pressure torque 55
- principal axes 13
- principia 7
- propagation 156
- quality factor 59, 94
- quasi-collisions 144
- Radau 104
- relationship
 - size-velocity 123
- relaxation modes 47
- resonance 163, 193

- angular variables 163
- ELR 193
- spin-orbit 159
- torque at 198
- resonant angle 175
- resonant returns 145
- rheological behavior 26
- rheology 101
- rheology equations 26
- ring
 - dissipative 186
 - flattening of 185
 - Jupiter, Uranus, Neptune 184
 - planetary 186, 192
 - resonances in planetary 192
 - Saturn's A, B and C 184, 191
 - thickness 190
- Roche zone 184
- rotation 2, 7, 23, 28
 - dynamics of 2, 3
 - long-axis mode 92
 - period 89, 109
 - short-axis mode 92
- rotational eigenmodes 57
- rotational perturbation 59
- rotational potential 54
- rubble pile 89, 94, 95, 99, 100, 133, 140
 - asteroids 95
- satellite 81
 - artificial 68
 - Galilean 163
 - orbit dynamic 82
- Satellite Laser Ranging 76
- scaling laws 118, 121
- Schiaparelli 50
- secular deceleration 59
- seismological data 26
- SELENE 85
- self-gravity 198
- SGG 84
- shear modulus 125
- simulations
 - of collisions 123
- size distribution 136, 137
- SLR 79, 83, 85
- solid inner core 50
- SPH 119, 130, 133
 - hydrocode 131
- spherical
 - inertia matrix 10
- spherical harmonics 33, 73, 84, 161
 - surface 75
- spin-orbit resonance 170
- SST 83
- STARLETTE 81
- strain
 - rate tensor 124
- strain tensor 32
- strength
 - compressive 128
 - dynamic 127
 - material 127
 - static 127
 - tensile 128
- stress 127
 - peak 127
 - peak failure 127
 - release 127
 - tensor 29, 102
- superball 9
- surface
 - equipotential 96
 - pressure 35
- Sylvia 89
- tensile strength 184
- tensor
 - stress 124
- thermodynamic 25
- tidal potential 54
- tidal torque 57, 59
- Tillotson
 - equation 125
 - equation of state 131
- tippy top 11
- Tokieda T. 1–22
- Toomre
 - parameter 189, 198
- top
 - Euler's 12, 17, 90
 - Kovalevskaya's 12, 19, 20
 - Lagrange's 12, 16, 17, 19
 - real 19
 - sleeping 19
 - spherical 14
- torque 6
 - satellite 199

- standard 199
- Toutatis 91, 95
 - asteroid 92
- tri-axial ellipsoids 97
- tumbling 92
- variables
 - action-angle 171
 - angles-actions 166
- velocity
 - angular 3, 10, 11
- Vesta 95
- viscoelastic body 44
- viscoelastic deformation 39
- viscoelastic model of Earth 59
- viscosity
 - effect of 197
- viscosity of the Earth 31
- volumic potential 35
- von Mises
 - criterion 125
- Voyager 191
- waves 198
 - elastic 124
 - plastic 124
 - shock 124
- Weibull
 - distribution 126, 129
 - parameters 126
- WKB
 - approximation 194
- wobble period 94
- Yarkovsky effect 136
- Young
 - modulus 101

Lecture Notes in Physics

For information about earlier volumes
please contact your bookseller or Springer
LNP Online archive: springerlink.com

- Vol.635: D. Alloin, W. Gieren (Eds.), Stellar Candles for the Extragalactic Distance Scale
- Vol.636: R. Livi, A. Vulpiani (Eds.), The Kolmogorov Legacy in Physics, A Century of Turbulence and Complexity
- Vol.637: I. Müller, P. Strehlow, Rubber and Rubber Balloons, Paradigms of Thermodynamics
- Vol.638: Y. Kosmann-Schwarzbach, B. Grammaticos, K. M. Tamizhmani (Eds.), Integrability of Nonlinear Systems
- Vol.639: G. Ripka, Dual Superconductor Models of Color Confinement
- Vol.640: M. Karttunen, I. Vattulainen, A. Lukkarinen (Eds.), Novel Methods in Soft Matter Simulations
- Vol.641: A. Lalazissis, P. Ring, D. Vretenar (Eds.), Extended Density Functionals in Nuclear Structure Physics
- Vol.642: W. Hergert, A. Ernst, M. Däne (Eds.), Computational Materials Science
- Vol.643: F. Strocchi, Symmetry Breaking
- Vol.644: B. Grammaticos, Y. Kosmann-Schwarzbach, T. Tamizhmani (Eds.) Discrete Integrable Systems
- Vol.645: U. Schollwöck, J. Richter, D. J. J. Farnell, R. F. Bishop (Eds.), Quantum Magnetism
- Vol.646: N. Bretón, J. L. Cervantes-Cota, M. Salgado (Eds.), The Early Universe and Observational Cosmology
- Vol.647: D. Blaschke, M. A. Ivanov, T. Mannel (Eds.), Heavy Quark Physics
- Vol.648: S. G. Karshenboim, E. Peik (Eds.), Astrophysics, Clocks and Fundamental Constants
- Vol.649: M. Paris, J. Rehacek (Eds.), Quantum State Estimation
- Vol.650: E. Ben-Naim, H. Frauenfelder, Z. Toroczkai (Eds.), Complex Networks
- Vol.651: J. S. Al-Khalili, E. Roeckl (Eds.), The Euroschool Lectures of Physics with Exotic Beams, Vol.I
- Vol.652: J. Arias, M. Lozano (Eds.), Exotic Nuclear Physics
- Vol.653: E. Papantonopoulos (Ed.), The Physics of the Early Universe
- Vol.654: G. Cassinelli, A. Leviero, E. de Vito, P. J. Lahti (Eds.), Theory and Application to the Galileo Group
- Vol.655: M. Shillor, M. Sofonea, J. J. Telega, Models and Analysis of Quasistatic Contact
- Vol.656: K. Scherer, H. Fichtner, B. Heber, U. Mall (Eds.), Space Weather
- Vol.657: J. Gemmer, M. Michel, G. Mahler (Eds.), Quantum Thermodynamics
- Vol.658: K. Busch, A. Powell, C. Röthig, G. Schön, J. Weissmüller (Eds.), Functional Nanostructures
- Vol.659: E. Bick, F. D. Steffen (Eds.), Topology and Geometry in Physics
- Vol.660: A. N. Gorban, I. V. Karlin, Invariant Manifolds for Physical and Chemical Kinetics
- Vol.661: N. Akhmediev, A. Ankiewicz (Eds.) Dissipative Solitons
- Vol.662: U. Carow-Watamura, Y. Maeda, S. Watamura (Eds.), Quantum Field Theory and Noncommutative Geometry
- Vol.663: A. Kalloniatis, D. Leinweber, A. Williams (Eds.), Lattice Hadron Physics
- Vol.664: R. Wiebeinski, R. Beck (Eds.), Cosmic Magnetic Fields
- Vol.665: V. Martinez (Ed.), Data Analysis in Cosmology
- Vol.666: D. Britz, Digital Simulation in Electrochemistry
- Vol.667: W. D. Heiss (Ed.), Quantum Dots: a Doorway to Nanoscale Physics
- Vol.668: H. Ocampo, S. Paycha, A. Vargas (Eds.), Geometric and Topological Methods for Quantum Field Theory
- Vol.669: G. Amelino-Camelia, J. Kowalski-Glikman (Eds.), Planck Scale Effects in Astrophysics and Cosmology
- Vol.670: A. Dinklage, G. Marx, T. Klinger, L. Schweikhard (Eds.), Plasma Physics
- Vol.671: J.-R. Chazottes, B. Fernandez (Eds.), Dynamics of Coupled Map Lattices and of Related Spatially Extended Systems
- Vol.672: R. Kh. Zeytounian, Topics in Hypersonic Flow Theory
- Vol.673: C. Bona, C. Palenzuela-Luque, Elements of Numerical Relativity
- Vol.674: A. G. Hunt, Percolation Theory for Flow in Porous Media
- Vol.675: M. Kröger, Models for Polymeric and Anisotropic Liquids
- Vol.676: I. Galanakis, P. H. Dederichs (Eds.), Half-metallic Alloys
- Vol.678: M. Donath, W. Nolting (Eds.), Local-Moment Ferromagnets
- Vol.679: A. Das, B. K. Chakrabarti (Eds.), Quantum Annealing and Related Optimization Methods
- Vol.680: G. Cuniberti, G. Fagas, K. Richter (Eds.), Introducing Molecular Electronics
- Vol.681: A. Llor, Statistical Hydrodynamic Models for Developed Mixing Instability Flows
- Vol.682: J. Souchay (Ed.), Dynamics of Extended Celestial Bodies and Rings