# Lecture Notes in Physics



Springer
*Berlin*
*Heidelberg*
*New York*
*Barcelona*
*Hong Kong*
*London*
*Milan*
*Paris*
*Tokyo*

## The Editorial Policy for Edited Volumes

The series *Lecture Notes in Physics* (LNP), founded in 1969, reports new developments in physics research and teaching - quickly, informally but with a high degree of quality. Manuscripts to be considered for publication are topical volumes consisting of a limited number of contributions, carefully edited and closely related to each other. Each contribution should contain at least partly original and previously unpublished material, be written in a clear, pedagogical style and aimed at a broader readership, especially graduate students and nonspecialist researchers wishing to familiarize themselves with the topic concerned. For this reason, traditional proceedings cannot be considered for this series though volumes to appear in this series are often based on material presented at conferences, workshops and schools.

## Acceptance

A project can only be accepted tentatively for publication, by both the editorial board and the publisher, following thorough examination of the material submitted. The book proposal sent to the publisher should consist at least of a preliminary table of contents outlining the structure of the book together with abstracts of all contributions to be included. Final acceptance is issued by the series editor in charge, in consultation with the publisher, only after receiving the complete manuscript. Final acceptance, possibly requiring minor corrections, usually follows the tentative acceptance unless the final manuscript differs significantly from expectations (project outline). In particular, the series editors are entitled to reject individual contributions if they do not meet the high quality standards of this series. The final manuscript must be ready to print, and should include both an informative introduction and a sufficiently detailed subject index.

## Contractual Aspects

Publication in LNP is free of charge. There is no formal contract, no royalties are paid, and no bulk orders are required, although special discounts are offered in this case. The volume editors receive jointly 30 free copies for their personal use and are entitled, as are the contributing authors, to purchase Springer books at a reduced rate. The publisher secures the copyright for each volume. As a rule, no reprints of individual contributions can be supplied.

## Manuscript Submission

The manuscript in its final and approved version must be submitted in ready to print form. The corresponding electronic source files are also required for the production process, in particular the online version. Technical assistance in compiling the final manuscript can be provided by the publisher's production editor(s), especially with regard to the publisher's own LaTeX macro package which has been specially designed for this series.

## LNP Homepage (http://www.springerlink.com/series/lnp/)

On the LNP homepage you will find:

—The LNP online archive. It contains the full texts (PDF) of all volumes published since 2000. Abstracts, table of contents and prefaces are accessible free of charge to everyone. Information about the availability of printed volumes can be obtained.

—The subscription information. The online archive is free of charge to all subscribers of the printed volumes.

—The editorial contacts, with respect to both scientific and technical matters.

—The author's / editor's instructions.

G. Contopoulos N. Voglis (Eds.)

# Galaxies and Chaos

Springer

**Editors**

George Contopoulos
and Nikos Voglis,
Academy of Athens,
Research Centre for Astronomy,
Anagnostopoulou Street 14,
106 73 Athens, Greece

Cataloging-in-Publication Data applied for

A catalog record for this book is available from the Library of Congress.

Bibliographic information published by Die Deutsche Bibliothek

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie; detailed bibliographic data is available in the Internet at http://dnb.ddb.de

ISSN 0075-8450
ISBN 3-540-40470-8 Springer-Verlag Berlin Heidelberg New York



Springer-Verlag Berlin Heidelberg New York
a member of BertelsmannSpringer Science+Business Media GmbH

http://www.springer.de



Typesetting: Camera-ready by the authors/editor
Camera-data conversion by Steingraeber Satztechnik GmbH Heidelberg
Cover design: *design & production*, Heidelberg

Printed on acid-free paper
54/3141/du - 5 4 3 2 1 0

# Preface

During the last two decades the science of nonlinear and chaotic dynamics has had a spectacular development. Many important ideas and tools appeared in the literature, helping to give a deeper understanding of the role of order and chaos in dynamical systems. One of the most fruitful applications of these ideas and tools has been in the field of dynamical astronomy, namely in galactic dynamics and in the dynamics of the solar system. On the other hand recent observational studies of galaxies and of exosolar systems have come to the point of detecting order and chaos in these systems. For this reason the members of the Research Center of Astronomy of the Academy of Athens decided to organize an international workshop on this subject. This workshop "Galaxies and Chaos. Theory and Observations" was held in Athens in September 16-19, 2002, (see http://www.cc.uoa.gr/gc2002/). A total number of 77 participants from 21 countries from all over the World attended the workshop, namely from Europe, U.S.A, Australia, Japan and Chile. There were 45 talks (23 of them invited talks) and 10 posters. The workshop brought together the experience of people working on galactic dynamics and galaxy formation (theory and observations) with the experience of people working on nonlinear dynamical systems. The talks summarized the most recent developments in both theoretical and observational aspects of galactic dynamics with emphasis on the role of chaos in galaxies. Studies of chaos in galaxies use methods similar to those frequently used in celestial mechanics, or other branches of physics and astronomy. For this reason we invited some speakers from related fields of research. A few interesting papers on some of the most up-to-date problems of celestial mechanics are included in this volume.

The Scientific Organizing Committee was composed of: G. Contopoulos (chairman, Academy of Athens), E. Athanassoula (Observatoire de Marseille, France), A. Bosma (Observatoire de Marseille, France), H. Dejonghe (University of Ghent, Belgium), A. Fridman (Russian Academy of Sciences), P. Grosbøl (ESO, Germany), P.O. Lindblad (Stockholm Observatory, Sweden), D. Lynden-Bell (University of Cambridge, UK), D. Merritt (Rutgers University, USA), and N. Voglis (Academy of Athens). The Local Organizing Committee was composed of: N. Voglis (chairman), H. Dara, Ch. Efthymiopoulos, P. Patsis, V. Tritakis and M. Zoulias.

The Academy of Athens covered a considerable part of the expenses of the workshop. But we are grateful also to several other institutions and persons, namely: The University of Athens, in particular the vice-rector Dr. G. Der-

mitzakis, that provided both financial and substructure support, the Hellenic Ministry of Culture, the A.G. Leventis Foundation, the City of Athens, Siemens S.A. in Athens, the European Physical Society and private donors. With their help we could in particular organize an archeological tour of Athens, a closing dinner at the terrace of a hotel in the center of the city, and provide free hotel rooms and free lunches to many participants. We thank heartily all of them.

The Editors
G. Contopoulos, N. Voglis

# Contents

## Part II Orbit Theory



## Part III Observations

# List of Contributors

**Athanassoula Lia**
Observatoire de Marseille,
France
lia@obmara.cnrs-mrs.fr

**Bacon Roland**
Centre de Recherche Astronomique de Lyon,
France
bacon@obs.univ-lyon1.fr

**Barkov Maxim**
Space Research Institute,
Russia
barmv@sai.msu.ru

**Barot J.M.**
Westminster School,
U.K.
jaideep.barot@westminster.org.uk

**Belinski Vladimir A.**
National Institute of Nuclearphysics (INFN) and International Center of Relativistic
Astrophysics (ICRA),
Italy
volodia@vxrmg9.icra.it

**Bergamin Jeroen**
University of Patras,
Greece

**Bērziņš Kārlis**
Ventspils International Radio Astronomy Center,
Latvia
kberzins@latnet.lv

**Bisnovatyi-Kogan Genadii**
Space Research Institute,
Russia
gkogan@mx.iki.rssi.ru

**Bertin Giuseppe**
Universit'a degli Studi di Milano,
Italy
bertin@sns.it

**Bosma Albert**
Observatoire de Marseille,
France
bosma@batis.cnrs-mrs.fr

**Bountis Tassos**
University of Patras,
Greece
bountis@math.upatras.gr

**Bureau Martin**
University of Columbia,
U.S.A.
bureau@astro.columbia.edu

**Burkert Andreas**
MPI Heidelberg
Germany,
burkert@mpia-hd.mpg.de

**Cappellari Michele**
Leiden University,
Netherlands
cappellari@strw.leidenuniv.nl

**Carati Andrea**
Universita di Milano,
Italy
carati@mat.unimi.it

**Ciotti Luca**
Universit'a di Bologna,
Italy
ciotti@bo.astro.it

**Contopoulos George**
Academy of Athens,
Research Center for Astronomy,
Greece
gcontop@cc.uoa.gr

**Copin Yannick**
Institut de Physique Nucl'eaire de Lyon,
France
y.copin@ipnl.in2p3.fr

**Davies Roger L.**
University of Oxford,
U.K.
rld@astro.ox.ac.uk

**De Bruyne Veronique**
University of Ghent,
Astronomical Observatory,
Belgium
Veronique.DeBruyne@rug.ac.be

**De Zeeuw Tim**
Leiden University,
Netherlands
tim@strw.leidenuniv.nl

**Dejonghe Herwig**
University of Ghent,
Astronomical Observatory,
Belgium
Herwig.Dejonghe@rug.ac.be

**Docenko Dmitrijs**
University of Latvia,
Latvia
dima@latnet.lv

**Dultzin-Hacyan Deborah**
Instituto de Astronomia de la Univ. Nacional Autonoma di Mexico,
Mexico
deborah@astroscu.unam.mx

**Dvorak Rudolf**
University of Vienna,
Austria
dvorak@astro.univie.ac.at

**Eckart Andreas**
University of Cologne,
Germany
eckart@ph1.uni-koeln.de

**Efthymiopoulos Christos**
Academy of Athens,
Research Center for Astronomy,
Greece
cefthim@cc.uoa.gr

**Eichler David**
Ben-Gurion University,
Israel
eichler@bgumail.bgu.ac.il

**Emsellem Eric**
Centre de Recherche Astronomique de Lyon,
France
emsellem@obs.univ-lyon1.fr

**Franco-Balderas Alfredo**
Instituto de Astronomia de la Univ. Nacional Autonoma di Mexico,
Mexico
alfred@astroscu.unam.mx

**Fridman Alexei**
Russian Academy of Sciences,
Russia
fxela@online.ru

**Galgani Luigi**
Universita di Milano,
Italy
galgani@berlioz.mat.unimi.it

**Gedalin Michael**
Ben-Gurion University,
Israel
gedalin@bgumail.bgu.ac.il

**Genzel Reinhard**
Max-Planck-Institut für extraterrestrische Physik,
Germany
genzel@imprs-astro.mpg.de

**Gottesman Stephen**
University of Florida,
U.S.A.
gott@astro.ufl.edu

**Griv Evgeny**
Ben-Gurion University,
Israel
griv@bgumail.bgu.ac.il

**Grosbøl Preben**
ESO,
Garching,
Germany
pgrosbol@eso.org

**Hadjidemetriou John**
University of Thessaloniki,
Greece
hadjidem@physics.auth.gr

**Harsoula Maria**
Academy of Athens,
Research Center for Astronomy,
Greece
mharsoul@phys.uoa.gr

**Hernández-Toledo Héctor**
Instituto de Astronomia de la Univ.
Nacional Autonoma di Mexico,
Mexico

**Hozumi Shunsuke**
Max-Planck-Institut fuer Astronomie,
Heidelberg,
Germany
hozumi@mpia-hd.mpg.de

**Hunter Chris**
Florida State University,
U.S.A.
hunter@math.fsu.edu

**Iyer Mansie**
Florida International University,
U.S.A.
miyer01@fiu.edu

**Kalapotharakos Costas**
University of Athens,
Greece
ckalapot@cc.uoa.gr

**Kandrup Henry Emil**
University of Florida,
U.S.A.
kandrup@astro.ufl.edu

**Khoruzhii Oleg**
Troitsk Institute for Innovation and
Thermonuclear Researches,
Russia
okhor@inasan.rssi.ru

**Kuntschner Harald**
European Southern Observatory,
Germany

**Lindblad Per A.B.**
Stockholm University Observatory,
Sweden

**Lindblad Per Olof**
Stockholm University Observatory,
Sweden
po@astro.su.se

**Liverts Edward**
Ben-Gurion University,
Israel
eliverts@bgumail.bgu.ac.il

**Lynden-Bell Donald**
University of Cambridge,
U.K.
dlb@ast.cam.ac.uk

**Maciejewski Witold**
Osservatorio Astrofisico di Arcetri,
Italy
witold@arcetri.astro.it

**Malphrus Benjamin**
Morehead State University,
U.S.A.
b.malphr@morehead-st.edu

**McDermid Richard**
Leiden University,
Netherlands
mcdermid@strw.leidenuniv.nl

**Miller Richard H.**
University of Chicago,
U.S.A.
rhm@oddjob.uchicago.edu

**Miller Bryan W.**
Gemini Observatory,
Chile
bmiller@gemini.edu

**Mond Michael**
Ben-Gurion University,
Israel
mond@menix.bgu.ac.il

**Mouawad Nelly**
University of Cologne,
Germany
nelly@ph1.uni-koeln.de

**Muccione Veruska**
Observatoire de Geneve,
Switzerland
veruska.muccione@obs.unige.ch

**Naab Thorsten**
Institute of Astronomy,
Cambridge,
UK
naab@ast.cam.ac.uk

**Neishtadt Anatoly**
Space Research Institute,
Russia
aneishta@vm1.iki.rssi.ru

**Nykytyuk Tetyana**
National Academy of Sciences of Ukraine, Ukraine,
nikita@mao.kiev.ua

**Ott Thomas**
Max-Planck-Institut für extraterrestrische Physik,
Germany
ott@mpe.mpg.de

**Peletier Reynier**
University of Nottingham,
U.K.
reynier.peletier@nottingham.ac.uk

**Pfalzner Susanne**
University of Cologne,
Germany
pfalzner@ph1.uni-koeln.de

**Polyachenko Evgenii**
Institute of Astronomy,
RAS,
Russia
epolyach@inasan.rssi.ru

**Psyhoyos Dionysia**
University of Thessaloniki,
Greece
dpsyc@skiathos.physics.auth.gr

**Sagdeev Roald**
University of Maryland,
U.S.A.
rzs@umd.edu

**Schödel Rainer**
Max-Planck-Institut für extraterrestrische Physik,
Germany
rainer@mpe.mpg.de

**Simpson Caroline**
Florida International University,
U.S.A.
simpsonc@fiu.edu

**Spurzem Rainer**
Astronomisches Recheninstitut,
Germany
spurzem@ari.uni-heidelberg.de

**Stavropoulos Ioannis**
University of Athens,
Greece
istavrop@cc.uoa.gr

**Straubmeier Christian**
University of Cologne,
Germany
cstraubm@ph1.uni-koeln.de

**Kleomenis Tsiganis**
Aristotle University of Thessaloniki,
Greece,
tsiganis@astro.auth.gr

**Van de Ven Glenn**
Leiden University,
Netherlands
vdven@strw.leidenuniv.nl

**Varvoglis Harry**
University of Thessaloniki,
Greece
varvogli@astro.auth.gr

**Verolme Ellen**
Leiden University,
Netherlands
verolme@strw.leidenuniv.nl

**Voglis Nikos**
Academy of Athens,
Research Center for Astronomy,
Greece
nvogl@cc.uoa.gr

**Voyatzis George**
University of Thessaloniki,
Greece
voyatzis@auth.gr

# Order and Chaos in Astronomy

George Contopoulos

Research Center of Astronomy, Academy of Athens, 14, Anagnostopoulou St., Athens, GR-10673, Greece

**Abstract.** We review the applications of order and chaos in various branches of astronomy. Order and chaos appear in generic dynamical systems, including the sun and other stars, the solar system and galaxies, up to the whole Universe. We discuss in particular the various types of orbits in galaxies, emphasizing the role of diffusion of chaotic orbits and the escapes to infinity. Then we consider chaos in dissipative systems, like gas in a galaxy, chaos in relativity and cosmology, and chaos in stellar pulsations and in the solar activity.

## 1 Introduction

The study of order and chaos had an explosive development in recent years. Thousands of papers were published on this subject. Particular problems of interest for Astronomy have been studied in various fields. Such fields are:

1. Celestial Mechanics
2. Galactic Dynamics
3. Relativity
4. Cosmology
5. Stellar Pulsations
6. Solar Activity

In the present review we will discuss several problems from these fields.

A special book by G. Contopoulos on "Order and Chaos in Dynamical Astronomy" (Springer Verlag, 2002) [20] has just appeared. This book of 624 pages and 305 figures has about 1200 selected references for further reading on this subject.

## 2 Celestial Mechanics

There are two very different traditions in Dynamical Astronomy. One deals mainly with regular phenomena (periodic and quasi-periodic motions) and the other with irregular phenomena (chaotic motions).

The basic example of order was provided by celestial mechanics. Strictly speaking only in integrable systems all motions are regular. But the systems considered in celestial mechanics are assumed to be close to integrable, and most motions are close to quasi-periodic. Thus the solar system was considered for a long time a paradigm of order.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 3–29, 2003.

Only in recent years chaos has been found in the solar system. The first chaotic phenomena referred to the irregular rotation of some satellites and to the distribution of the asteroids ([81], [48], [82], [46], [40], [65]). On the other hand the motions of the planets seem to be stable over several billions of years. But chaos is present in the motions of the planets also. The time scale for chaos (the so-called Lyapunov time) is relatively short in the case of Pluto (some $10^7$ years). Nevertheless Pluto does not escape from the solar system, because of some resonance effects. Thus Pluto is an example of "stable chaos".

A similar phenomenon of "stable chaos" appears also in the case of certain asteroids that fill some resonances (and do not form gaps there), despite their short Lyapunov time [63]. A new explanation of this phenomenon will be given by H. Varvoglis during this workshop.

Another planet with Lyapunov time shorter than a Hubble time is Mercury. According to Laskar [55] Mercury's orbit should approach eccentricity 1 after about 3.5 billion years, therefore Mercury should either plunge into the sun, or collide with Venus. However, R. Dvorak, will present numerical evidence in this workshop that Mercury will be stable for several Hubble times.

There are two more phenomena that affect our Earth, and deal with order and chaos. One is the origin of the meteorites that hit the Earth. These meteorites and the dust of the zodiacal light come from very different parts of the solar system, and their orbits are in general chaotic [34].

The other phenomenon deals with the obliquity of the Earth's axis, which is stabilized by the action of the Moon [54]. Without the Moon the Earth's orientation would change completely, in a chaotic way, and this should affect considerably the evolution of life. However, with the Moon in its present position the chaotic variation of the Earth's obliquity does not exceed certain limits, and the corresponding climatic changes are also limited.

## 3 Classification of Chaotic Systems

Chaotic behaviour is a general characteristic of nonintegrable dynamical systems. There are various degrees of chaos that are classified in classical books of statistical mechanics on regular and chaotic motion (e.g. [58]). Namely nonintegrable systems are considered to be ergodic, mixing, Kolmogorov, or Anosov (each class is a subclass of the previous class). They are ergodic if most motions go everywhere in the available phase space, mixing if two nearby particles deviate considerably, Kolmogorov if this deviation is exponential in time, and Anosov if the average exponential factor (the so-called Lyapunov characteristic number) is always larger than a nonzero number. In random systems we have the limit of an infinite Lyapunov characteristic number.

However, this classification is now obsolete. In fact ergodic systems (and mixing, Kolmogorov and Anosov systems) appear only rarely, and generic dynamical systems contain both order and chaos. Namely the set of ordered motions is not zero in general, and this is the opposite of what happens in ergodic systems. This conclusion is based on the famous KAM theorem ([53], [1], [67]), that proves

**OLD CLASSIFICATION**

Integrable systems
Ergodic systems
Mixing systems
Kolmogorov systems
Anosov systems

**NEW CLASSIFICATION**

| | **ORDERED** | **CHAOTIC** | | |
|---|---|---|---|---|
| **COMPACT** | Integrable | (General Case) Systems with Divided Phase Space | (Limiting Cases) Ergodic Mixing Kolmogorov Anosov | **RANDOM** |
| **NONCOMPACT** | Integrable with escapes | Nonintegrable with escapes (Chaotic scattering) | — | |

**Fig. 1.** Old and New Classification of Dynamical Systems.

the existence of a finite set of quasi-periodic motions in generic dynamical systems. Thus the presently accepted classification of dynamical systems is shown in Fig. 1.

We call chaotic systems those having chaotic domains, but possibly also ordered domains. In limiting cases we have only chaotic (ergodic) motions.

A system is close to integrable if its chaotic domains are small and it is close to ergodic if its ordered domains are small. The random systems are extremely chaotic. Their Lyapunov time is zero. As regards the noncompact systems (i.e. systems with escapes) they may be integrable or nonintegrable. In the latter case we have the phenomenon of "*chaotic scattering*".

A general theorem that has been proved recently in the case of some simple maps, namely the logistic map [51] and the standard map [36] is the following. If the perturbation parameter K is large there are many values of K for which there are no stable periodic orbits and the system is completely chaotic (i.e. the ordered orbits have a measure zero). Nevertheless there is no interval $\varDelta$K of values of K without any islands of stability. This theorem seems to be applicable in generic dynamical systems.

Examples of Hamiltonian systems with such a behaviour have been found in recent years [25] and [44]. These systems have a central periodic orbit which is alternatively stable and unstable, up to arbitrarily large values of the perturbation.

A classical case where we have both order and chaos is the standard map

$$\begin{aligned} x' &= x + y' \\ &\quad (mod 1) \\ y' &= y + \frac{K}{2\pi} sin(2x) \end{aligned} \tag{1}$$

For small K (Fig. 2a) this system is mostly ordered, with only a little chaos near the unstable periodic orbit (0,0) and the asymptotic curves emanating from it.

As K increases chaos increases also (Figs. 2b,c,d,e) and for K=8 it seems that chaos is complete.

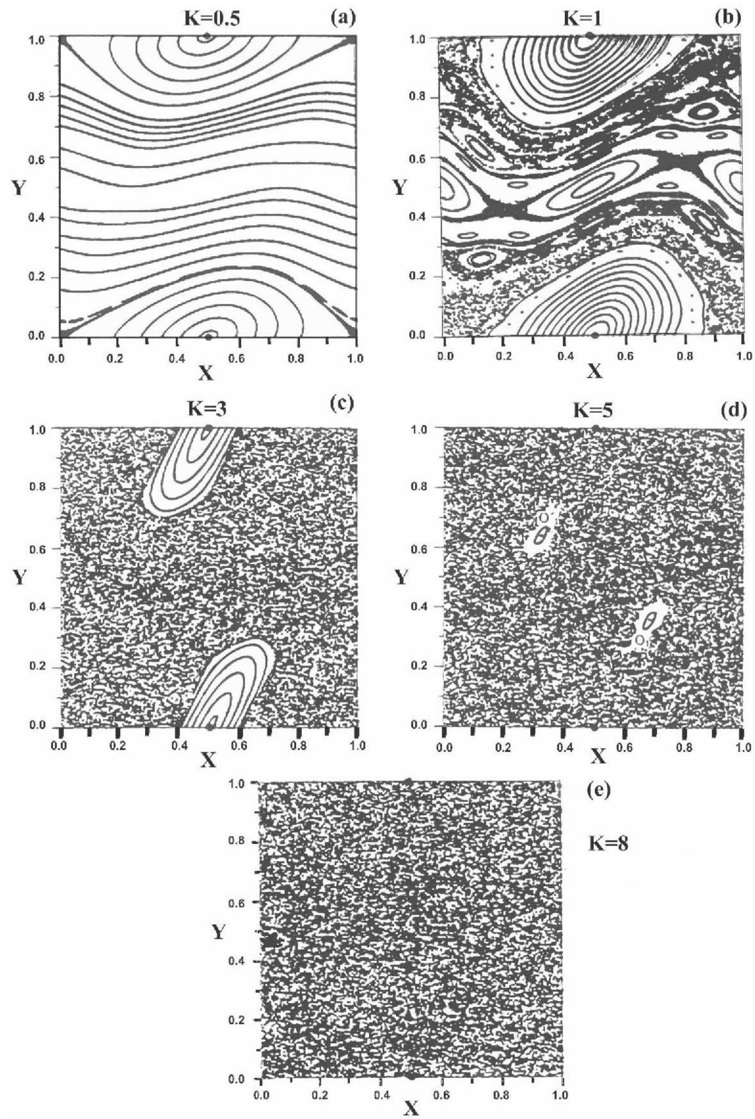



**Fig. 2.** Phase portraits of the standard map for various values of the nonlinearity parameter K: (a)K=0.5, (b)K=l, (c)K=3, (d)K=5, (e)K=8.

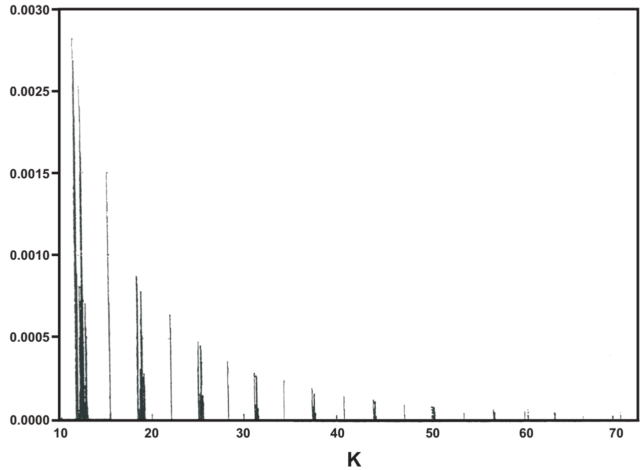



**Fig. 3.** The area of the islands of stability in the standard map as a function of K [with unit the square (0,l)x(0,l)].

Nevertheless small islands of order have been found for larger values of K in a recurrent way (Fig. 3). Namely islands appear in particular regions of the phase space whenever K increases by $2\pi$ [37].

Therefore chaos is never complete, despite a widespread opinion that beyond a limiting perturbation the system should remain completely chaotic.

## 4 Galactic Dynamics

Order and chaos play an important role in galactic dynamics. The appearance of order in galactic dynamics is based on the "third integral" of motion [14].

One of the first applications of the third integral was in the velocity ellipsoid of stars near the sun. If there are only two integrals of motion, the energy

$$E = \frac{1}{2}(R^2 + Z^2 + \Theta^2) + V \tag{2}$$

and the angular momentum

$$J = r\Theta \tag{3}$$

one should have a distribution function of the form

$$f = f(E, J) \tag{4}$$

In particular an ellipsoidal distribution should be of the form

$$f = f(R^2 + Z^2 + k(\Theta - \Theta_0)^2 \tag{5}$$

with the two equal axes along R and Z. However, the observations had shown definitely that the Z axis is much shorter than R. This indicated the existence of a third integral of motion

$$I = Z^2 + \dots \tag{6}$$

that should be included in the distribution function f for the velocity [5], so that the distribution function should become

$$f = f(R^2 + (1 + k')Z^2 + k(\Theta - \Theta_0)^2) \tag{7}$$

My involvement in the subject of the third integral in galaxies started during my first visit to Stockholm in 1956. I worked with Professor Bertil Lindblad on a generalization of the epicylic theory of planar galactic orbits. But when I tried to extend the theory to three dimensions I could not do very much analytically. At that time Per Olof Lindblad was calculating planar orbits in galaxies to explain the spiral arms. His first calculations were presented at an IAU Symposium in 1956 [59] At my request he calculated for me two orbits in three dimensions. The surprising thing was that these orbits (Fig. 4) were not ergodic, as I expected, but indicated the existence of a third integral of motion in generic dynamical systems ([13],[14]). I realized later that Whittaker [80] had already found a third integral (that he called adelphic integral) in particular cases by a different method.

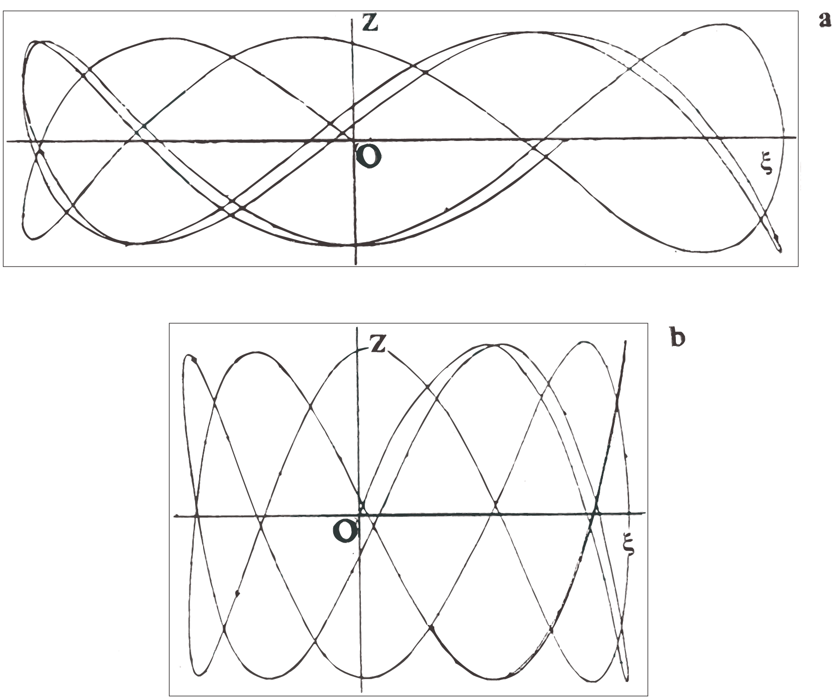



**Fig. 4.** The first calculated orbits in the meridian plane of an axisymmetric galaxy (1956) are like deformed Lissajous figures.

It is remarkable that Birkhoff [7] had also found a similar integral by another different method, but he never believed in its usefulness. In fact the third integral is in general not exact, but only a formal series. According to Birkhoff this indicated that such an integral would be applicable only over limited times. In particular he believed that a linearly stable periodic orbit in a system of two degrees of freedom would be stable for all times only in integrable systems.

It was only through the KAM theorem that the complete stability of linear stable orbits was established for generic two-dimensional dynamical systems. But even after the KAM theorem was established, it was not clear what was the usefulness of such a formal integral.

I remember that when I first spoke to Moser about the third integral in galactic dynamics he was doubtful of its usefulness, as it is only an asymptotic series that does not converge. But at that time I could calculate by computer algebra ([21], [16]) the higher order terms of the third integral (this was done by simple Fortran, without the present packages of computer algebra, like mathematica etc). The results were remarkable, indicating an excellent agreement between theoretical and numerical orbits when higher order terms of the third integral were calculated. This impressed Moser and he remarked in one of his papers ([68]) that because of this work in galactic dynamics the subject of the third integral "received renewed interest".

At the same time a student of Arnold, Nekhoroshev, studied the usefulness of truncated third integrals. He found the best truncation, and he showed the applicability of these integrals over exponentially long times, a result that is the central element of the Nekhoroshev [69] theory.

In some cases one has to reach very high orders of truncation of the third integral in order to find the Nekhoroshev limit and deviations beyond it.

Another independent development started with the Fermi-Pasta-Ulam paper [39] on coupled oscillators that revealed the existence of ordered orbits in systems of many degrees of freedom. This result was also explained, later, by means of formal integrals of motion, of the third integral type.

## 5 Chaotic Orbits in Galaxies

A recent example refers to an application of the third integral to self-consistent models of galaxies, generated by N-body simulations [28]

Up to now in most applications the potential of the galaxy was assumed to be given by a simple analytic formula. Then a third integral could be constructed in order to find the structure of the regular orbits. But in our present studies we start with models produced by the collapse of a protogalaxy. This gives not only the density and the potential of the final model, after the collapse, but also the distribution of the velocities. This method is very different from other methods, like the Schwarzschild (1979) method, that try to construct self-consistent models by populating various initial conditions of orbits in an appropriate way.

The success of the original Schwarzschild method, indicated that most orbits are regular, i.e. his self-consistent models were very close to integrable. At the same time several integrable models, the so-called Stäckel models, were introduced in galactic dynamics ([35], [8], [78], [50], etc). In particular, several self-consistent models were constructed. This led to the conjecture that chaos is unimportant in galactic dynamics. It was stated loosely that somehow, Nature avoids chaos in galaxies and forms only integrable models.

However, our studies of models generated by N-body simulations indicate that in many cases both ordered and chaotic orbits are present. There are three main types of orbits in a galaxy, box orbits (Fig. 5a), tube orbits (Fig. 5b) and chaotic orbits (Fig. 5c). Box orbits appear in nonrotating systems (e.g. elliptical galaxies). Tube orbits appear near various resonances (see also Figs. 6a,b,c). Chaotic orbits are due to an interaction of resonances. E.g. in Fig. 7 a chaotic orbit joins the bar region of a galaxy with the outer near circular orbits. Chaos is small in some cases, but not negligible. On the other hand no model was found that is completely chaotic.

The general conclusion is that Nature does not form integrable models like those derived from Stäckel potentials, nor ergodic systems, like those used in statistical mechanics, except in an approximate way.

There are three types of chaotic orbits in galaxies:

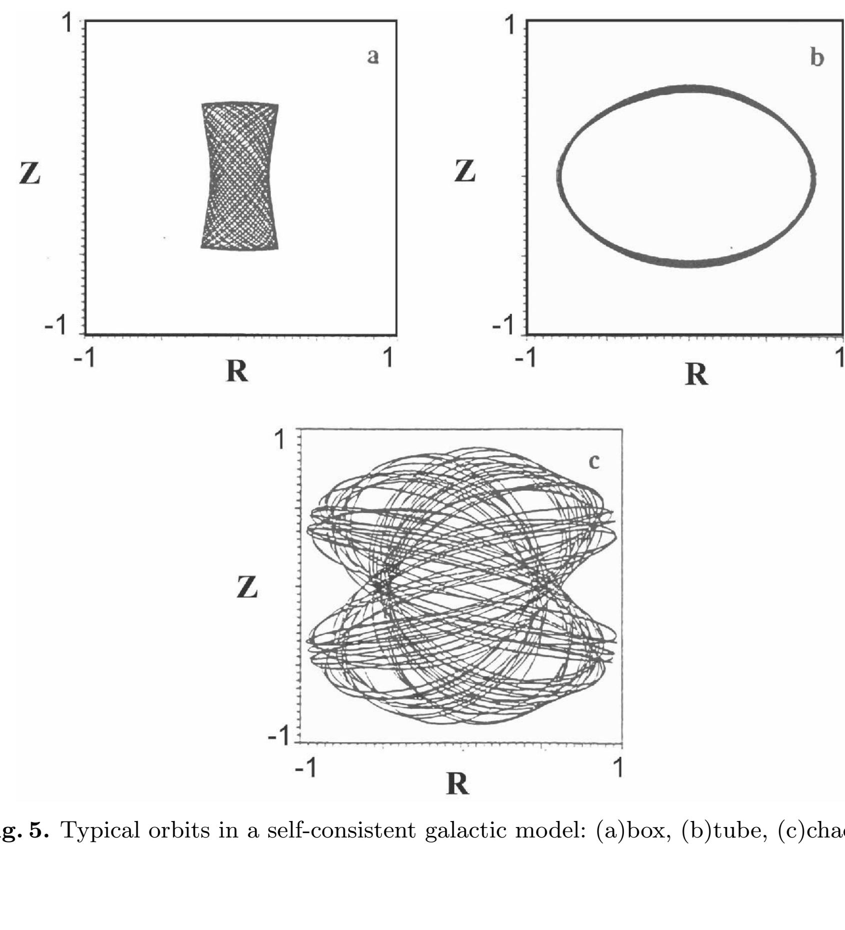



**Fig. 5.** Typical orbits in a self-consistent galactic model: (a)box, (b)tube, (c)chaotic.

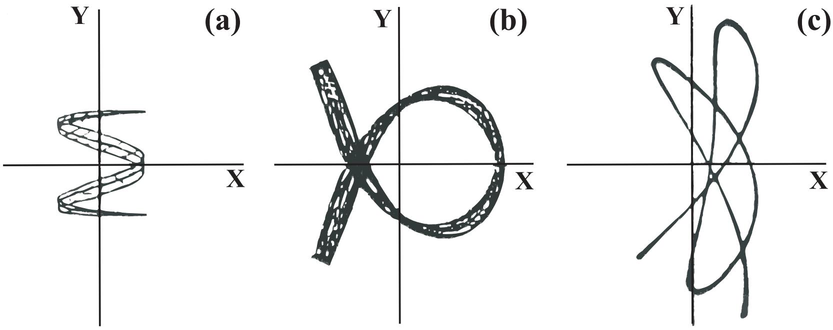



**Fig. 6.** Three further types of tube orbits.

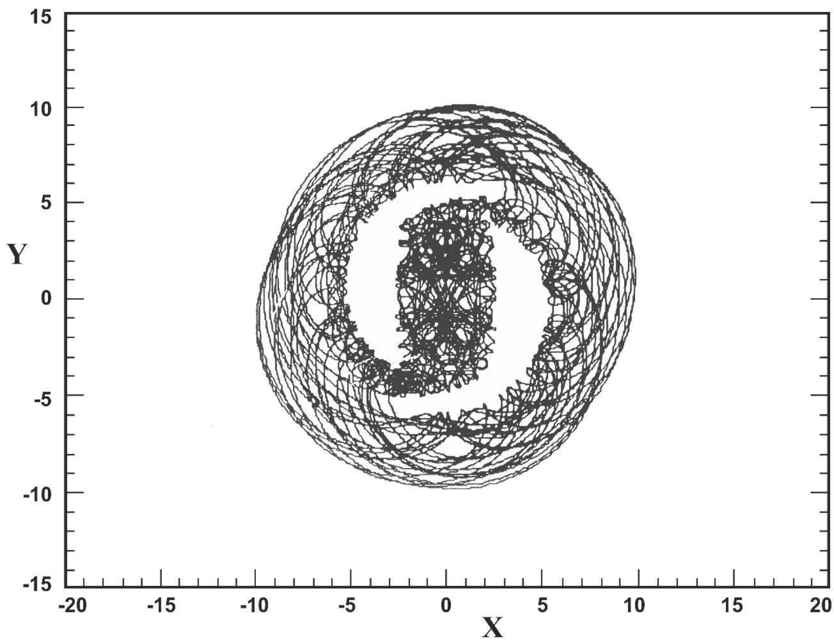



**Fig. 7.** A chaotic orbit in a barred galaxy.

1. Orbits near corotation in rotating spiral, or barred galaxies [17]. Such orbits are important in terminating the bars close to corotation.
2. Elongated orbits passing near the center of a galaxy ([61], [43], [62]). Such orbits are particularly important when there is a strong mass concentration near the center, e.g. a black hole. A large black hole in a nonspherical galaxy tends to change the character of the orbits, from boxes filling curvilinear parallelograms far the center, to almost Keplerian orbits near the center. The mixing of two types of orbital behaviour along the same orbit produces chaos. One can even estimate the size of a black hole by the degree of chaos produced around it.

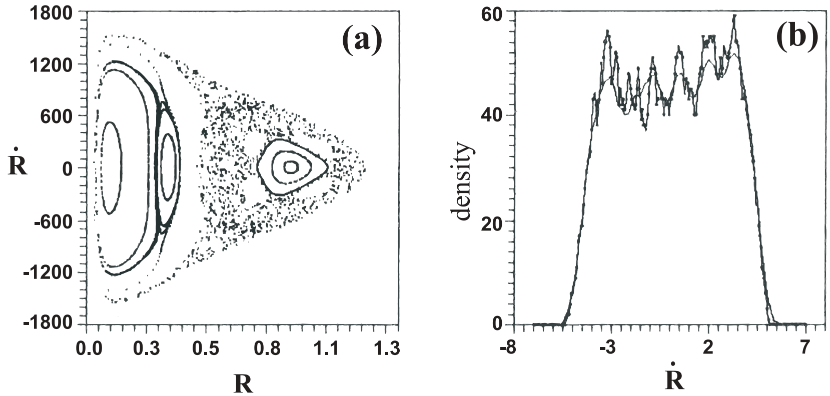



**Fig. 8.** (a) Ordered and chaotic orbits on a surface of section in a self-consistent galactic model. (b) The density in the chaotic domain (R=0.6) is roughly constant.

3. A third chaotic region is the region between the box orbits, that cover a large inner part of a galaxy (except the very central region) and the tube orbits that refer mainly to orbits circulating around the main galaxy in the outer parts (Fig. 8). We notice that chaotic orbit produce a constant density in the chaotic domain.

## 6 Diffusion of Galactic Orbits

When an integrable system is slightly perturbed it develops a small degree of chaos. This is introduced as follows. In an integrable case the Poincaré surface of section is filled with invariant curves. We may consider a general case in which the invariant curves close around a central invariant point O. The rotation number along successive invariant curves varies smoothly. For every rational rotation number n/m all the orbits starting on the corresponding invariant curve are periodic, of period m (Figs. 9a,b). When a generic perturbation of order $\epsilon$ is introduced, only two periodic orbits of period m are left, one stable and one unstable (Fig. 9c). (In some cases there are more pairs of stable-unstable orbits). The stable orbits are surrounded by islands, while near the unstable points some chaos appears.

As the perturbation increases the islands increase in size. The chaotic regions also increase in size forming zones of instability. However between the main zones of instability there are still invariant curves, around the center,and the various chaotic regions are separated.

When the perturbation $\epsilon$ goes beyond a critical value $\epsilon_{crit}$ the various chaotic regions communicate (Fig. 9c) and chaos becomes important abruptly. This is the phenomenon of "resonance overlap" ([16], [74], [84], [11]). This leads to a diffusion of the orbits that is called "resonance overlap diffusion". This is the main mechanism that introduces a large degree of chaos.

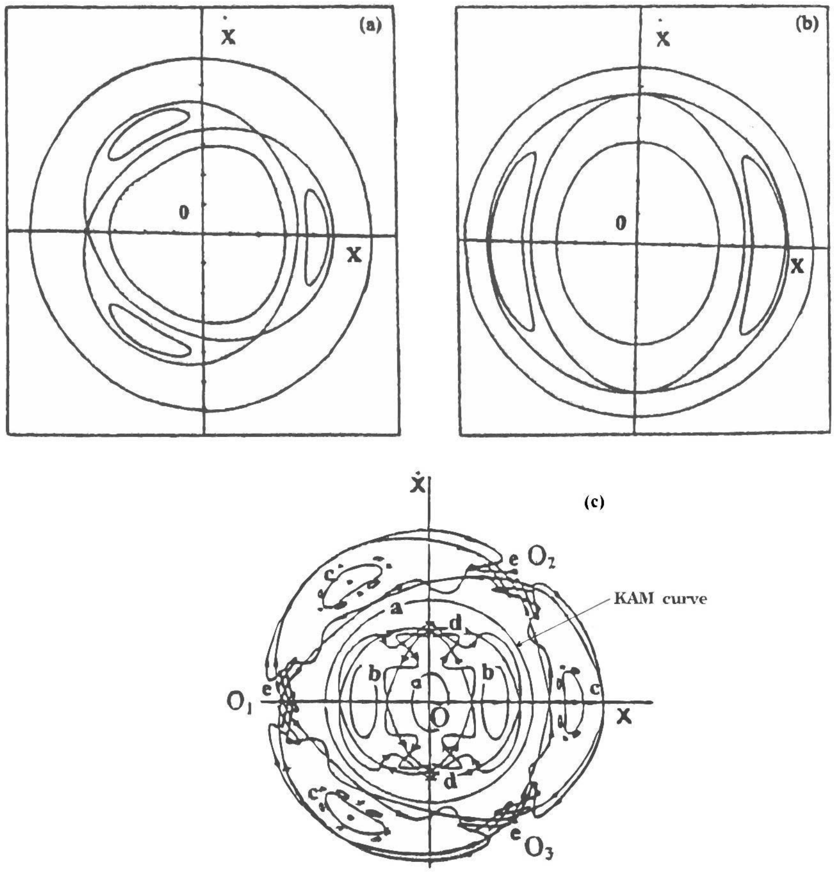



**Fig. 9.** Invariant curves on a surface of section of two resonant integrable cases and one nonintegrable case. (a) Case with three islands, (b) Case with two islands. (c) A case with both double and triple islands. For small perturbations the chaotic domains near the unstable double and triple orbits are separated by KAM curves. But for larger perturbations the last KAM curve is destroyed and large chaos, due to resonance overlap, is produced.

In three or more degrees of freedom there is one more mechanism that introduces chaos, namely Arnold diffusion [3]. In fact, while in systems of two degrees of freedom the KAM surfaces are two-dimensional and separate the interior from the exterior in the 3-D phase space, in three degrees of freedom the KAM surfaces are three-dimensional and do not separate the phase space, which is now five-dimensional. Thus the various chaotic regions always overlap, even for arbitrarily small perturbation $\epsilon$.

However the time scale of Arnold diffusion is very large, of order

$$T \propto exp(\frac{1}{\epsilon}) \tag{8}$$

i.e. it is exponentially long ([3], [69], [11]). In the close neighbourhood of invariant tori the time scale is even superexponential, of the form

$$T \propto exp(exp(\frac{1}{\epsilon})) \tag{9}$$

or even longer [66]. In galactic problems this time is much longer than the age of the Universe, therefore Arnold diffusion is not important. Only in plasma physics this diffusion may have observable consequences.

Although Arnold diffusion is also due to resonance overlap, nevertheless we distinguish clearly the resonance overlap diffusion from Arnold diffusion ([54], [22]) even in systems of three or more degrees of freedom. In fact, Arnold diffusion appears only very close to the resonance lines of the Arnold web (Fig. 10) while resonance overlap extends over large regions of the phase space.

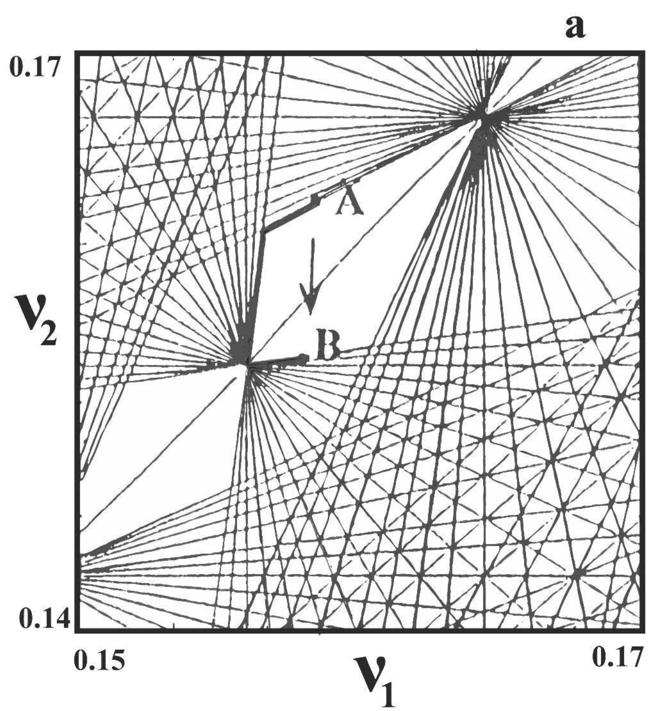



**Fig. 10.** The Arnold web contains the resonant lines between the frequencies $\nu_1$ and $\nu_2$. If the perturbation is small there is a slow Arnold diffusion from A to B along the thick lines. But if the perturbation is large the resonant lines become very thick and allow resonance overlap diffusion, directly from A to B.

A numerical example was given in the case of two coupled standard maps [22]

$$x_1{}' = x_1 + y_1{}' \quad , \quad y_1{}' = y_1 + \frac{K}{2\pi} sin2\pi x_1 - \frac{\beta}{\pi} sin2\pi(x_2 - x_1\ )$$
$$(\ mod1) \quad (10)$$
$$x_2{}' = x_2 + y_2{}' \quad , \quad y_2{}' = y_2 + \frac{K}{2\pi} sin2\pi x_2 - \frac{\beta}{\pi} sin2\pi(x_1 - x_2\ )$$

where K is the nonlinearity and $\beta$ the coupling parameter.

The diffusion time T is given empirically by the formula

$$T = exp[a - b(\beta - \beta_{crit})] \tag{11}$$

where K=3, a=4.14, b=4160, and $\beta_{crit} = 0.305124$ is a critical value of the coupling parameter (Fig. 10).

For $\beta > \beta_{crit}$ the diffusion time decreases with $\beta$ according to the exponential law (11) (Fig. 11) and becomes very small for relatively large $\beta$. However, for ($\beta < \beta_{crit}$ the time T increases superexponentially as $\beta$ decreases and for $\beta < 0.305$ it becomes larger than T=$10^{10}$ iterations, i.e. it is very difficult to calculate numerically.

In systems of two degrees of freedom there is also a slow diffusion, like Arnold diffusion, when the orbits have to cross the holes of cantori, surrounding islands of stability.

Cantori appear near the outermost invariant curve surrounding a stable invariant point, when the perturbation K increases beyond a critical value $K_{crit}$. Then the outermost invariant curve is destroyed and it becomes a cantorus with

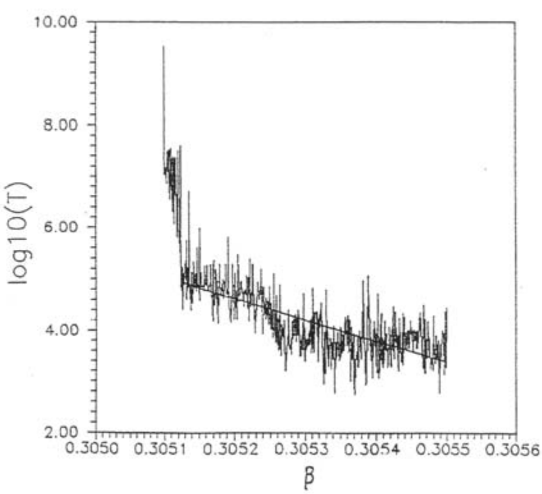



**Fig. 11.** The diffusion time T as a function of the coupling parameter $\beta$ increases exponentially for decreasing $\beta$, if $\beta > 0.305124$, and superexponentially for smaller $\beta$.

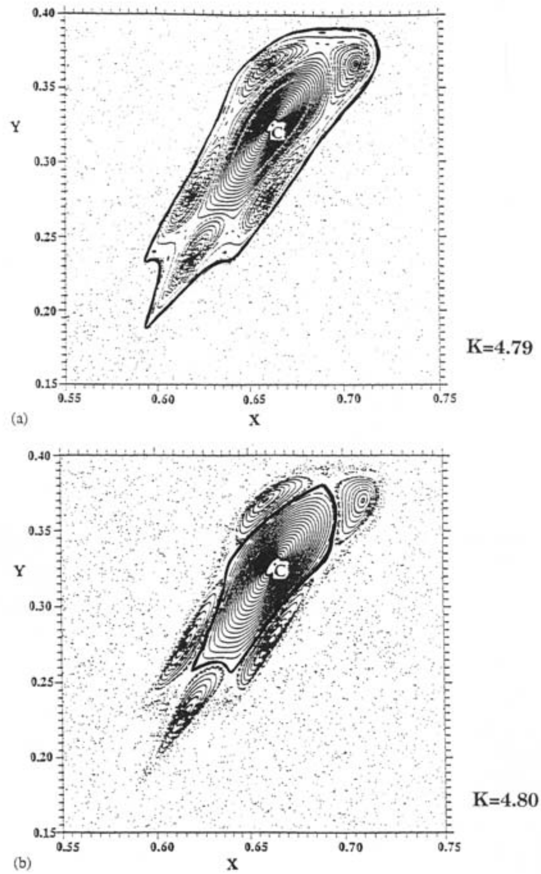



**Fig. 12.** An island of stability in the standard map inside the large chaotic sea is limited by the last KAM curve, (a) For K=4.79 the last KAM curve surrounds 5 islands of stability, (b) for K=4.80 the last KAM curve is inside these islands.

infinite holes. Chaotic orbits close to this cantorus can cross it from inside outwards, or from outside inwards. Thus we see two different chaotic domains outside the main island of stability. A "sticky zone" between the island and the cantorus and a "large chaotic sea" outside the cantorus (Fig. 12). When K goes beyond $K_{crit}$ the size of the island decreases abruptly and a sticky zone is formed between the cantorus and the new outermost limit of the island. The diffusion time increases exponentially as we deviate from the cantorus inwards and approach the new boundary of the island.

On the other hand if the perturbation increases, the holes of the cantorus become larger and the diffusion time becomes abruptly much shorter.

As an example we consider the cantori around an important island in the standard map (see (1)) for K=5 and K=4.998. In the first case we found orbits that escape to the chaotic sea through the cantorus in a time T=2, while in the second case this time increases to T=2xl0$^5$ [28].

Despite these variations the stickiness effect plays a role in practical applications. For example in galactic dynamics some orbits stay close to the boundary of a resonant island for a long time, and can be considered as regular during the life-time of the galaxy, although in the long run they may become very chaotic and may even escape from the system.

## 7 Escapes

Another interesting subject related to chaos, deals with escapes. If the energy h of a star in a galaxy is larger than the escape energy $h_{esc}$, the star may escape to infinity. However, there are many stars, with energy larger than $h_{esc}$, which do not escape, because they are trapped around a stable periodic orbit.

One way to study the problem of escapes is by using a simple model. An example is given by the Hamiltonian

$$H = \frac{1}{2}(\dot{x}^2 + \dot{y}^2 + Ax^2 + By^2) - \epsilon xy^2 = h \tag{12}$$

(A=1.6, B=0.9, $\epsilon = 0.08$ and various values of the energy h). The escape energy in this case is $h_{esc}$=25.31.

If his larger than $h_{esc}$ the CZV has two openings (Fig. 13). At the openings, there are two unstable periodic orbits, that are called Lyapunov orbits. Any particle crossing a Lyapunov orbit outwards escapes from the system and never returns to it.

For an energy slightly smaller than the escape energy most initial conditions generate chaotic orbits. Most of these orbits escape when the energy increases slightly above the escape energy. However, the escape may take a long time. There are sets of orbits that escape rather fast, after one or two iterations, but there are also orbits that escape only after many iterations. The domains of escape are limited by orbits that are asymptotic to one or the other Lyapunov orbits.

In Fig. 15 we show the escape domains for orbits escaping after one or two iterations in the forward time direction from the upper opening only.

By counting the escaping particles out of a large number of bodies we found empirically the escape rate and the number of remaining particles N out of an initial number $N_0$ [24]

$$N = N_0 e^{-pt} \tag{13}$$

If the energy is larger than a critical value $h_{crit}$, much larger than escape energy $h_{esc}$, practically all particles escape according to the law (13) and the escape rate p is proportional to a power of (h-$h_{crit}$),

$$p \propto (h - h_{crit})^2 \tag{14}$$

where a=0.5, practically the same for several potentials. Thus the exponent a may be a universal number [52].

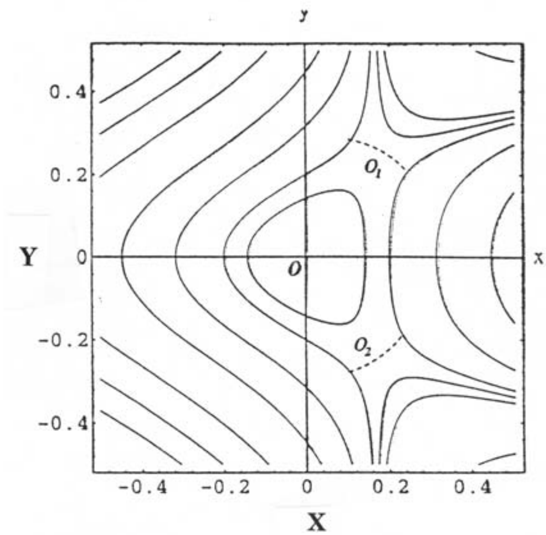



**Fig. 13.** Equipotentials in the Hamiltonian (12). These are open if the energy is larger than the energy of escape. At the openings, there are two unstable periodic orbits, $O_1$ and $O_2$, called Lyapunov orbits.

If the energy is larger than $h_{esc}$, but smaller than $h_{crit}$, there is a set of particles $N_{non}$, that never escape. Then the escape law is of the form

$$N - N_{non} = (N_0 - N_{non})e^{-pt} \tag{15}$$

The transition at $h_{crit}$ is like a phase transition and it is connected with the size of the islands of stability. For $h < h_{crit}$ there are important islands of stability, while for $h > h_{crit}$ the size of the islands becomes very small, and most islands disappear altogether.

The fact that some particles escape from the system has as a consequence the nonconservation of areas on a surface of section (Fig. 14). Namely, while for $h < h_{esc}$ the plane y=0 is a Poincaré surface of section and the areas are conserved, for $h > h_{esc}$ this surface is no more a Poincaré surface of section and the areas are reduced.

Thus for $h > h_{esc}$ the 2-D map on the surface y=0 looks like a dissipative system although in the full 3-D space the volumes are preserved. Furthermore, although conservative systems do not have attractors, nevertheless in systems with escapes the infinity acts as an attractor. And if there are more than one openings of the curves of zero velocity we may speak of a corresponding number of co-existing attractors.

## 8 Dissipative Systems

The main difference between dissipative systems and conservative systems is the existence of attractors at a finite distance.

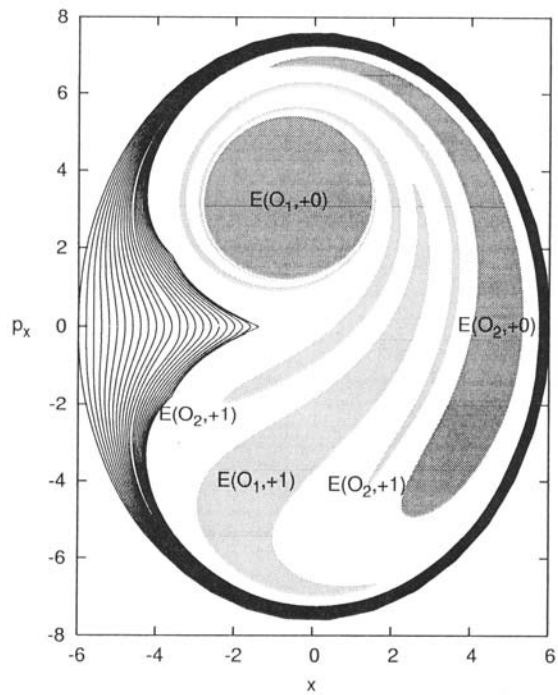



**Fig. 14.** A (non Poincaré) surface of section for the Hamiltonian (12) and an energy larger than the escape energy. The orbits starting in the domains $E(O_1,n)$ escape through the Lyapunov orbit $O_1$ after n further crossings with the surface of section. Near the boundary there are invariant curves of nonescaping orbits.

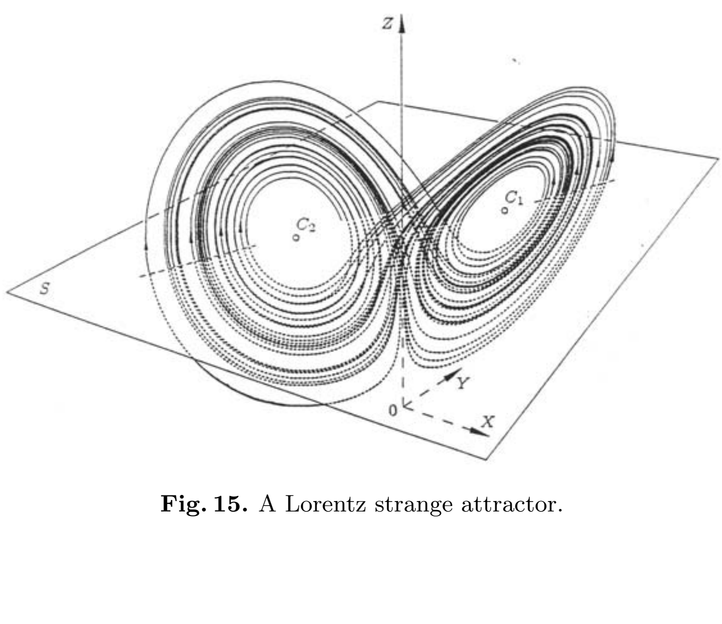



**Fig. 15.** A Lorentz strange attractor.

In general, dissipative systems have three types of attractors (1) Point attractors, (2) Limit cycles (in two dimensional flows) and (3) strange attractors. In more dimensions we may have limiting structures of higher than one dimension.

We have two types of dissipative systems (1) those given by nonlinear differential equation, and (2) those given by nonlinear maps.

A simple dissipative system of differential equations is the Lorentz system [60]

$$\begin{aligned} dx/dt &= \sigma(y-x) \\ dy/dt &= \rho x - y - xz \\ dz/dt &= -\beta z + xy \end{aligned} \tag{16}$$

(where $\sigma$, $\rho$, $\beta > 0$).

A most simple dissipative map is the Hénon map [47]

$$x' = y - Kx^2 + 1 \quad , \quad y' = bx \tag{17}$$

with $0 < b < l$.

These systems were the first to provide strange attractors (Fig. 15).

In dissipative cases the volume, or surface, in phase space is shrinking. The Lyapunov characteristic number is negative in the cases of point attractors or limit cycles, but it is positive in the case of strange 'attractor. Namely, the moving points approach the attractor, but nearby points on the attractor deviate exponentially. The Hénon map (17) has a limiting conservative case for b=l. In this case no attractor appears.

There are some classical books on chaos in dissipative systems, like the books of Lichtenberg and Lieberman [58] and Guckenheimer and Holmes [45].

In galaxies the gas is a dissipative system. Thus, we have a secular evolution of the gas that is different from the behaviour of the stars. In particular we may have an attractor at the center of the galaxy, or a limit cycle in the case of a stationary flow of gas. A special case is the appearance of vortices near corotation (Fig. 16) in spiral and barred galaxies ([38], [41], [42]), a subject that will be discussed by Fridman during this workshop.

## 9 Chaos in Relativity and Cosmology

Much work has been done on this subject in recent year. An example of this activity is the volume on "Deterministic Chaos in General Relativity" edited by Hobill et al. [49]. A more recent review was provided by Contopoulos et al. [27].

One of the first cases where chaos was found in General Relativity was the case of two fixed black holes ([18], [19]). It is remarkable that the relativistic problem is chaotic, while the corresponding classical problem of two fixed centers is completely integrable. In particular the relativistic motions of photons are completely chaotic, but the phase space extends to infinity. On the other hand the motions of particles with nonzero rest mass are in part ordered and in part chaotic.

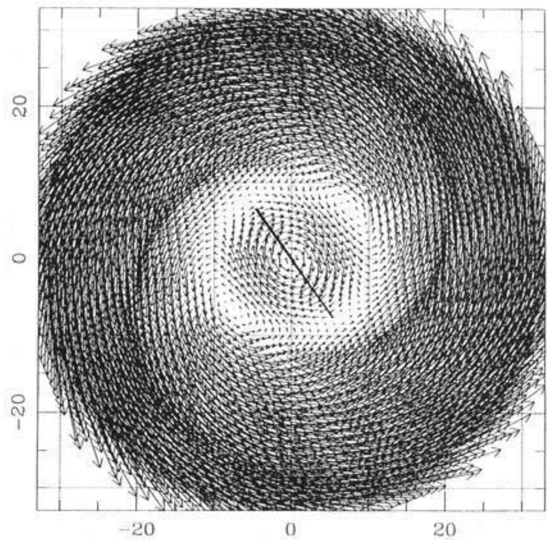



**Fig. 16.** The flow of gas in a barred galaxy. In the corotation region there are two cyclones along the bar and two anticyclones perpendicularly to the bar.

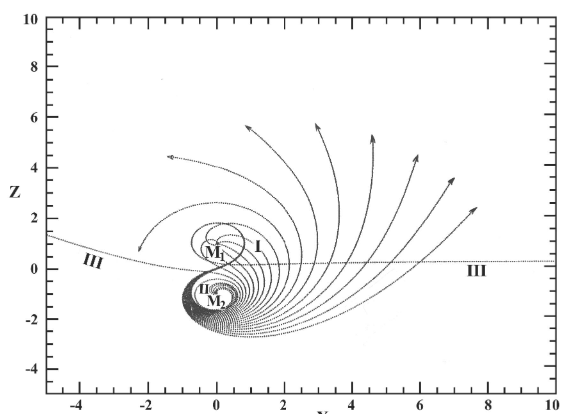



**Fig. 17.** Chaos in the case of two fixed black holes. Orbits of photons coming from infinity fall into the black hole $M_1$ (orbits of type I), or $M_2$ (orbits of type II), or escape to infinity (orbits of type III).

In the first case the photons may escape to one of the two black holes, or to infinity. A beam of photons (Fig. 17) coming from infinity is separated into three sets, one (I) leading to the black hole $M_1$ the second (II) leading to the black hole $M_2$, and the third (III) leading to infinity.

The three sets are fractal and their basic property is that between two orbits of different sets there is one orbit of the third set. For example in Fig. 17 between the orbits 1 leading to $M_1$ and the orbit 2 leading to infinity, there is an orbit

3 leading to $M_2$ (in fact there is a set of orbits (III) leading to $M_2$). Similarly between the orbit 1 ($M_1$) and the orbit 3($M_2$) there is the orbit 4 leading to infinity, etc.

The three sets of orbits I, II and III are intertwined in a fractal way.

In the case of particles of nonzero rest mass, with energy smaller than the escape energy, the orbits can escape to the black holes $M_1$ and $M_2$ only. In such a case there are some stable periodic orbits and orbits close to them do not escape to any of the black holes.

In recent years much more work has been done in this problem and in other relativistic problems where chaos is present.

An important case with application to cosmology is the Mixmaster Universe model ([6], [64]). This represents a particular solution of Einstein's equations that has three different scale factors $\alpha, \beta, \gamma$ along the axes x, y, z. These equations are

$$\begin{aligned} 2\alpha &= (e^{2\beta} - e^{2\gamma})^2 - e^{4\alpha} \\ 2\beta &= (e^{2\gamma} - e^{2\alpha})^2 - e^{4\beta} \\ 2\gamma &= (e^{2\alpha} - e^{2\beta})^2 - e^{4\gamma} \end{aligned} \tag{18}$$

The basic property of this model is that the two scale factors are positive and one negative (or two negative and one positive). Therefore, the model expands along certain directions and contracts along others. But the directions of expansion and contraction change in a chaotic way.

There has been much analytical and numerical work on this model. A strange feature of this model is that the Lyapunov characteristic number is zero, yet the system seems to be chaotic. For some time it was expected that the Mixmaster model may be integrable, but more recently it was established that it is nonintegrable [57], [26], [31]. On the other hand the Mixmaster model is not ergodic [33], and has no periodic orbits.

These somewhat conflicting properties can be understood if we notice that the Mixmaster model is not compact, i.e. its orbits escape in general to infinity. In particular this property explains why the Lyapunov characteristic number is zero. In fact, if a particle tends to infinity, a nearby particle deviates from it linearly in time, i.e.

$$\xi = \xi_0 t \tag{19}$$

as in the case of two nearby orbits escaping from a galaxy. Therefore the Lyapunov characteristic number is given by

$$LCN = \lim_{t\to\infty} \frac{ln|\frac{\xi}{\xi_0}|}{t} = \lim_{t\to\infty} \frac{lnt}{t} = 0 \tag{20}$$

and this is equal to zero. But the finite time LCN is positive, and this indicates that the Mixmaster model is chaotic for all finite times (Fig. 18). This behaviour is similar to a chaotic scattering case [26].

There are more general cosmological models where chaos is important. In fact the role of chaos in Cosmology seems to be more and more appreciated in recent years.

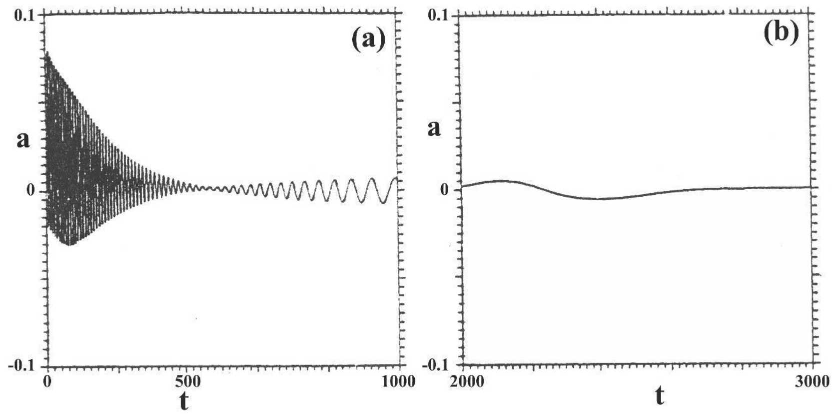



**Fig. 18.** The short time Lyapunov characteristic number in the mixmaster model as a function of t for (a) 0¡t¡1000, and (b) 2000¡t¡3000. The average value is positive and tends to zero as t.

In a similar way in the modern unified theories of the Universe (superstrings, quantum gravity, etc) chaos seems to play a significant role that only recently has started to be explored.

## 10 Chaos in Stellar Pulsations

The variable stars have only rarely completely regular pulsations. In most cases they have irregularities in the period and in the amplitude of the pulsations.

In many cases it is even impossible to define an average period of the variable stars. Various types of irregular variables appear in specific regions of the H-R diagram.

A review of the various studies of irregular variable stars was given some years ago by Perdang [70].

A classical method to study the variations in stellar models is numerical hydrodynamics. In the case of radial oscillations one separates the variable star in a number of concentric shells and follows their evolution in time [12] by solving the hydrodynamic and heat flow equations (Fig. 19). These solutions give good agreement with observations.

An alternative method is to consider the equations for the amplitudes of the most important excited modes. By solving numerically the so-called "amplitude equations" one finds results that are consistent with the hydrodynamic results [32], [9], [10].

The amplitude equations are of the form

$$\frac{d\alpha_1}{dt} = \kappa_i \alpha_i + N_i \tag{21}$$

where $N_i$ are functions containing the non-linear terms. What is remarkable is that accurate results are found if one considers only two or three modes (i=2, or 3), and $N_i$ contains only quadratic and cubic nonlinearities.

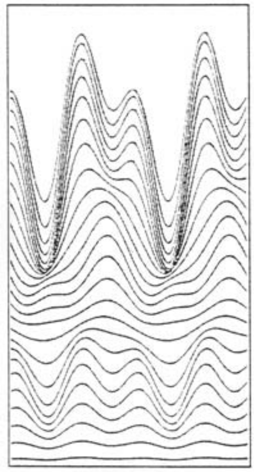



**Fig. 19.** The variations of the radii of various layers of a pulsating variable star as a function of t.

Thus, instead of the partial differential equations of hydrodynamics, we reach a problem of coupled ordinary differential equation, which is very similar to the problems of dissipative systems of particle dynamics, that we mentioned above.

A special case of particular interest is the case of radial perturbations around a state of hydrodynamical equilibrium. As it was pointed out by Woltjer [83] long ago, this problem can be formulated as a Hamiltonian system of a finite number N of degrees of freedom.

The Hamiltonian takes the form

$$H = \frac{1}{2}\sum_{i=1}^{N} {p_i}^2 + V \tag{22}$$

where V contains terms of order 2,3, and possibly higher order terms

$$V = \frac{1}{2}\sum_{i=1}^{N} {\omega_i}^2 {q_i}^2 + \frac{1}{3}\sum_{i,j,k=1}^{N} {V_{ijk}}^{(3)} q_i q_j q_k + ... \tag{23}$$

This Hamiltonian is of the same form, as those used in galactic dynamics. An extensive use of such a Hamiltonian in studying stellar pulsations was made by Perdang and Blacher [71], [72].

We will not discuss this subject further, but only refer to a related review article by Perdang [70]. Perdang and his associates found chaotic oscillations that can be described by such a Hamiltonian formalism. The analogy with our dynamical problem of galactic astronomy is astonishing.

The reduction of the partial differential equations of hydrodynamics (like those encountered in stellar pulsations) to a set of ordinary differential equations can be extended to more general problems. In fact there is an analogy between partial and ordinary differential equations. There are integrable and nonintegrable systems of partial differential equations, as there are integrable and nonintegrable systems of ordinary differential equations. In the integrable cases there are infinite conserved quantities, like the integrals of motion of the ordinary differential equations. These systems have as particular solutions the solitons, i.e. solitary waves that move unchanged, even after interacting nonlinearly with each other.

Of even greater importance is the fact that many nonintegrable systems of partial differential equations are close to integrable and they have soliton solutions, which decay only slowly in time.

This subject has a close analogy with the nonintegrable systems of ordinary differential equations that are close to integrable, like the systems encountered in galactic dynamics.

## 11 Chaos in Solar Activity

It is well known that the solar cycle repeats itself every 11 years, but not exactly in the same way (Fig. 20). This phenomenon can be explained by assuming that the solar cycle is due to a strange attractor ([76], [77]). This can be shown by reconstructing the sunspot data $M_i$ for the successive days i in a 3-dimensional space ($M_i$, $M_{i+d}$, $M_{i+rd}$), where d is a fixed delay time. A projection of the reconstructed manifold on a plane is shown in Fig. 21a. Spiegel and Wolf [77] selected the delay interval d=1200 days, after some experimentation. In Fig. 21a

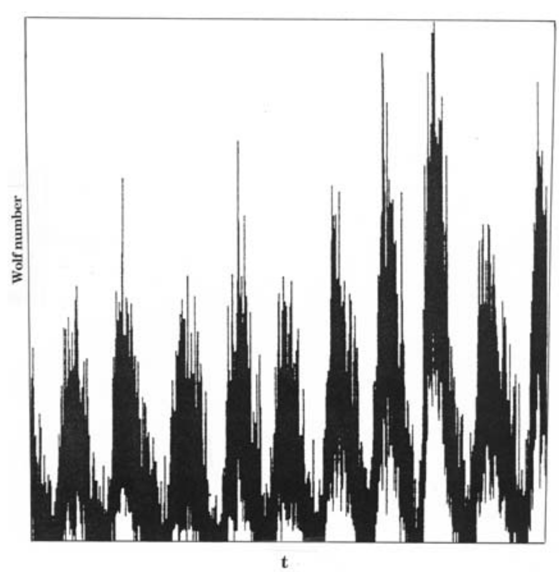



**Fig. 20.** The solar activity (Wolf numbers) follows the 11-year cycle, but different cycles are not the same.

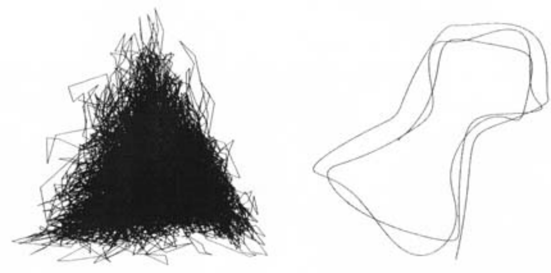



**Fig. 21.** Reconstruction of the sunspot data (Wolf numbers $M_i$, $M_{i+d}$, $M_{i+2d}$ for all days i and a delay d=1200 days), (a) Projection of the lines joining successive points on a plane, (b) After filtering the Fourier components by a cutoff at 1 year the figure tends to a strange attractor.

we see a chaotic behaviour, but no detailed information can be drawn from this figure. However, after using a filter that eliminates all Fourier components with period one year or less, one finds the curve of Fig. 21b, which is reminiscent of the Lorentz strange attractor of Fig. 15.

This strange attractor explains the variations of the sunspot cycles, and even more important variations like the Maunder minimum. This refers to the fact that the solar activity was practically nonexistent during the reign of Louis IV, the king that chose the sun as his emblem. During that period the sun was not covered by sunspots. This phenomenon can now be explained by means of a strange attractor with a 5-dimensional fractal dimension [77].

Therefore, the theory of chaos has important applications in explaining the solar activity. However, the work done in this field is still rather limited.

Another application refers to the solar flares. This phenomenon is a manifestation of an important chaotic process that is called "self-organized criticality" ([58],[4]). Namely the energy of the magnetic field is continuously accumulated in certain regions and it is released at irregular time intervals in the form of flares. This phenomenon is very similar to earthquakes, that also accumulate energy from the motion of the plates, and release it at irregular time intervals. This subject will be discussed during our workshop by Dr. Papazachos.

Therefore, the theory of chaos has useful applications in the most important solar phenomena. Similar phenomena appear in other types of stars, and in active galaxies, like active galactic nuclei and quasars.

However, a much more detailed study of these phenomena is required before we understand them completely. I only stress here the similarity of these problems with the corresponding phenomena of galactic dynamics from a mathematical point of view.

A conclusion from my review is that Order and Chaos play a very important role in Astronomy. Much work has been done already but the prospects of this new field are practically unlimited. I do not doubt that the study of Order and

Chaos will help to unify Physics and Astronomy from their most elementary constituents to the whole Universe.

## References


1. Arnold, V.I.: Sov. Math. Dokl. **2**, 245 (1961)
2. Arnold, V.I.: Sov. Math. Dokl. **5**, 581 (1964)
3. Arnold, V.I. and Avez, A.: *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968)
4. Bak, P.:*How Nature Works* (Springer Verlag, New York, 1996)
5. Barbanis, B.: Z. Astrophys. **56**, 56 (1962)
6. Belinskii, V.A. and Khalatnikov, I.M.: Sov. Phys. JETP **29**, 911 (1969)
7. Birkhoff, G.D.: *Dynamical Systems* (Amer.Math.Soc., Providence, R.I., 1927)
8. Bishop, J.L.: Astrophys. J. **305**, 14 (1986)
9. Buchler, J.R.: in *Chaos in Astrophysics*, ed. by Buchler, J.R., Perdang, J. and Spiegel, E.A. (Reidel, Dordrecht, 1985) p. 11
10. Buchler, J.R.: in *Chaotic Phenomena in Astrophysics*, ed. by Buchler,J.R and Eichhorn, H., N.Y. Acad.Sci. Annals **497**, 37 (1987)
11. Chirikov, B.V.: Phys.Rep. **52**, 263 (1979)
12. Christy, F.R. : Quart. R.Astr. Soc. **9**, 13 (1968)
13. Contopoulos, G.: Stockholm Ann. **20**, No 5 (1958)
14. Contopoulos, G.: Z. Astrophys. **49**, 273 (1960)
15. Contopoulos, G.: Astrophys. J. Suppl. **13**, 503 (1966)
16. Contopoulos, G.: in *Les Nouvelles Methodes de la Dynamique Stellaire*, ed. by Hénon M. and Nahon, F. Bull.Astron. (3) **2**, 223 (Besancon, CNRS, 1967)
17. Contopoulos, G.: Astron.Astrophys. **117**, 89 (1983)
18. Contopoulos, G.: Proc. Roy. Soc. London A**431**, 183 (1990)
19. Contopoulos, G.: Proc. Roy. Soc. London A**435**, 551 (1991)
20. Contopoulos, G.: *Order and Chaos in Dynamical Astronomy* (Springer Verlag, New York, 2002)
21. Contopoulos, G. and Moutsoulas, M.: Astron.J. **70**, 817 (1965)
22. Contopoulos, G. and Voglis, N.: Cel. Mech.Dyn.Astron. **64**, 1 (1996)
23. Contopoulos, G., Galgani, L. and Giorgilli, A.: Phys.Rev. A **18**, 1183 (1978)
24. Contopoulos, G., Kandrup, H.E. and Kaufmann, D.: Physica D **64**, 310 (1993)
25. Contopoulos, G., Papadaki, H. and Polymilis, C. : Cel.Mech.Dyn.Astron. **60**, 249 (1994)
26. Contopoulos, G. Grammaticos, B. and Ramani, A.: J. Phys. A **28**, 5313 (1995)
27. Contopoulos, G., Voglis, N. and Efthymiopoulos, C.: Cel.Mech.Dyn.Astr. **73**, 1 (1999)
28. Contopoulos, G., Efthymiopoulos, C. and Voglis, N.: Cel.Mech.Dyn.Astron. **78**, 243 (2000)
29. Contopoulos, G., Harsoula, M. and Voglis, N.:Cel.Mech.Dyn.Astron. **78**,197 (2000)
30. Contopoulos, G., Voglis, N. and Kalapotharakos, C.: Cel.Mech.Dyn.Astr. **83**, 191 (2002)
31. Cornish, N.J. and Levin, J.J.: Phys.Rev. D **55**, 7489 (1998)
32. Coullet, F. and Spiegel, E.A.: SIAM J.Appl.Math. **43**, 776 (1983)
33. Cushman, R. and Sniatycki, J.: Rep.Math.Phys. **36**,75 (1995)
34. Dermott, S.F. and Nicholson, P.D.: Highlights of Astronomy **8**, 259 (1989)

35. de Zeeuw, T.: Mon.Not.R.Astr.Soc. **216**, 273 (1985)
36. Duarte, P.: Ann.Inst.Henri Poincaré **11**, 359 (1994)
37. Dvorak, R., Freistetter, F., Funk, B. and Contopoulos G., in *Proceedings of the 3rd Austrian-Hungarian Workshop on Trojans and related topics*, ed. by Freistetter, F., Dvorak, R. and Erdi, B. (2003)
38. England, M.N., Hunter,J.H. Jr. and Contopoulos, G.: Astrophys. J. **540**, 154 (2000)
39. Fermi, E., Pasta, J. and Ulam, S.: Los Alamos Lab. Rep. LA 1949 (1955)
40. Ferraz-Mello, S.: in *Asteroids, Comets, Meteors 1993* ed. by Milani, A., di Martino, M. and Cellino, A. (Kluwer, Dordrecht, 1994) p. 175
41. Fridman, A.M., Khoruzhii, O.V., Polyachenko, V.L., Zasov, A.V., Silchenko, O.K., Moiseev, A.V., Burlak, A.N., Afanasiev, V.L., Dodonov, S.N. and Knapen, J.H.: Mon. Not R. Astr. Soc. **323**, 651 (2001)
42. Fridman, A.M., Khoruzhii, O.V., Lyakhovich, V.V., V.L., Silchenko, O.K.,Zasov, A.V., Afanasiev, V.L., Dodonov, S.N. and Boulesteix, J.: Astron. Astrophys. **771**, 538 (2001)
43. Gerhard, O.E. and Binney, J.: Mon.Not.R.Astr.Soc. **216**, 467 (1985)
44. Grousousakou, E. and Contopoulos, G.: in *The Dynamics of Small Bodies in the Solar System* ed. by Steves, B.A. and Roy, A.E. (Kluwer, Dordrecht, 1999) p. 535
45. Guckenheimer, J. and Holmes, P.: *Nonlinear Oscillations, Dynamical Systems and Bifurcations of Vector Fields* (Springer Verlag, N.York, 1983)
46. Hadjidemetriou, J.: Cel.Mech.Dyn.Astron. **56**, 563 (1993)
47. Hénon, M.: Comm.Math.Phys. **50**, 69 (1976)
48. Henrard, J.: in *Long Term Behaviour of Natural and Artificial N-Body Systems*, ed. by Roy A.E. (Kluwer, Dordrecht, 1983) p. 405
49. Hobill, D., Burd,A. and Coley, A. (eds): *Deterministic Chaos in General Relativity* (Plenum Press, New York, 1994)
50. Hunter, C. and de Zeeuw, P.T.: Astrophys. J. **389**, 79 (1992)
51. Jacobson, M.V.: Commun.Math.Phys. **81**, 39 (1981)
52. Kandrup, H.E., Siopis, C., Contopoulos, G. and Dvorak, R.: Chaos **9**, 381 (1999)
53. Kolmogorov, A.N.: Dokl. Akad. Nauk. SSSR **98**, 527 (1954)
54. Laskar, J. : Physica D **67**, 257 (1993)
55. Laskar, J. : Astron.Astrophys. **287**, L9 (1994)
56. Laskar, J., Joutel, F. and Robutel, P.: Nature **361**, 615 (1993)
57. Latifi, A., Musette, M. and Conte, R.: Phys. Lett. A **194**, 83 (1994)
58. Lichtenberg, A.J and Lieberman, M.A. : *Regular and Chaotic Dynamics*, 2nd Ed. (Springer Verlag, New York, 1992)
59. Lindblad, B. and Lindblad, P.O.: in *IAU Symposium 5*, 8 (1958)
60. Lorentz, E.N.: J.Atmos. Sci. **20**, 130 (1963)
61. Mayer, F. and Martinet, L.: Astron.Astrophys. **27**, 199 (1973)
62. Merritt, D. and Fridman, T.: Astrophys.J. **460**, 136 (1996)
63. Milani, A. and Nobili, A.M.: Nature **357**, 569 (1992)
64. Misner, C.M.: Phys. Rev. Lett. **22**, 1071 (1969)
65. Moons, M. and Morbidelli, A.: Icarus **114**, 33 (1995)
66. Morbidelli, A. and Giorgilli, A.: J. Stat. Phys. **78**, 1607 (1995)
67. Moser, J. : Nachr. Acad. Wiss. Gottingen II. Math. Phys. Kl. 1 (1962)
68. Moser, J. : Mem. Amer. Math. Soc. **81**, 1 (1968)
69. Nekhoroshev, N.N. : Russ. Math. Surv. **32**(6), 1 (1977)
70. Perdang, J.: in "Chaos in Astrophysics" ed. by Buchler,J.R, Perdang, J. and Spiegel, E.A., (Reidel, Dordrecht, 1985) p. 11

71. Perdang, J. and Blacher, S.: Astron.Astrophys. **112**, 35 (1982)
72. Perdang, J. and Blacher, S.: Astron.Astrophys. **136**, 263 (1984)
73. Pettini, M. and Vulpiani, A.: Phys.Lett.A **106**, 207 (1984)
74. Rosenbluth, M.N., Sagdeev, R.A., Taylor, J.B. and Zaslavsky, G.M.: Nucl. Fusion **6**, 217 (1966)
75. Schwarzschild, M.: Astrophys. J. **232**, 236 (1979)
76. Spiegel, E.A.: in Spiegel, E.A. and Zahn, J-P. (eds), "Lecture Notes in Physics", **71**, 3 Springer Verlag, New York (1977)
77. Spiegel, E.A. and Wolf, A.: in *Chaotic Phenomena in Astrophysics* ed. by Buchler,J.R and Eichhorn, H., N.Y. Acad.Sci. Annals **497**, 55, (1987)
78. Statler, T.S.: Astrophys. J. **321**, 113 (1987)
79. Sussman, G.J. and Wisdom, J.: Science **241**, 433 (1988)
80. Whittaker, E.T.: *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies*, 4th Ed., (Cambridge Univ. Press, Cambridge, 1937)
81. Wisdom, J.: Astron. J. **87**,577 (1982)
82. Wisdom, J., Peale, S.J. and Mignard, F.: Icarus **58**,137 (1984)
83. Woltjer, J.: Mon.Not.R.Astr.Soc. **95**,260 (1935)
84. Zaslavsky, G.M. and Chirikov, B.V.: Sov.Phys.Uspekhi **14**, 549 (1972)

# Critical Ergos Curves and Chaos at Corotation

Donald Lynden-Bell[1,2,3] and Jaideep M. Barot[1,4]

[1] Institute of Astronomy, The Observatories, Cambridge, CB3 0HA.
[2] Clare College, dlb@ast.cam.ac.uk.
[3] Physics Dept, The Queen's University, Belfast, BT7 1NN.
[4] Westminster School, jaideep.barot@westminster.org.uk.

**Abstract.** The theory of adiabatic invariants is developed to cover the gyration of a star about a nearly equipotential orbit in a galaxy with a strong bar. The guiding centres for such orbits follow curves of constant Ergos. The energy and the gyration adiabatic invariant give two constants of the motion. Critical Ergos curves have a pair of $X$-type gravitational neutral points which provide switches between trajectories that have the star circulating forward or backward relative to the corotating frame of the bar and those that liberate back to remain on one side of the galaxy's centre.

An attempt to discover the dynamical basis of the apparently random switching, that has been observed in computations of orbits with finite amplitudes of gyration, FAILS to find any such chaos at small gyration amplitudes, where Ergos curves give a good description of guiding centre motion.

## 1 Introduction

Eddington [15] looked for solutions of the collisionless Boltzmann equation that lacked axial symmetry but were steady in non-rotating axes. He introduced the idea of principal velocity surfaces to which the principal axes of the velocity ellipsoid were orthogonal. He then proved that, if the velocity ellipsoids were triaxial corresponding to three independent integrals quadratic in the velocities, the principal velocity surfaces had to be confocal quadrics. Also the potential had to be of a special form corresponding to Stackle's separable systems. Chandrasekhar [6] vehemently criticised Eddington's assumption that principal velocity surfaces existed but the analysis without that assumption produced no new solutions of interest. Meanwhile Clarke [7] derived the algebraic integrals corresponding to Eddington's system which were exploited to great effect by Kuzmin [20] and others. Lynden-Bell [21] gave a new analysis without assuming that the integrals were quadratic, but while he derived all six integrals and showed that the turning points lay on the confocal quadrics, he again found no new systems. It was de Zeeuw [12] & [13] who's careful categorisations of the orbital structure in these separable systems that revived interest in them. For an elementary derivation in axial symmetry, see Lynden-Bell [25]. Rather less is known about the analytic form of the integrals of the motion in systems that are only steady when viewed in rotating axes. Freeman [16] gave a fine analysis of the special systems in which the forces are linear functions of position which form a natural development of Riemann's homogeneous ellipsoids. Vandervoort [28] discovered a Stackle system in rotating axes which was further developed by Contopoulos

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 30–43, 2003.

and Vandervoort [9] but the density corresponding to this special potential is not positive everywhere. de Zeeuw and Merritt [11] developed a theory suitable for the cores of rotating systems, while Berman and Mark [3] analysed nearly circular orbits trapped in weakly non-linear spiral waves and gave analytical approximations to the slightly non-circular motion of the guiding centres. Binney & Tremaine [4] gave a general discussion of computed orbits.

For individual orbits a significant advance was made by L.S. Hall [17] who asked for invariant relations for one energy rather than an integral of the motion for all energies. This gave him a far wider class of potentials than those for which exact integrals exist Whittaker [29] Marshall & Wojciechowski [26].

Here we develop the adiabatically invariant gyration of a star about a guiding centre to give us an approximate integral independent of the energy, which is especially useful in the complicated region of barred galaxies close to corotation. The analysis of orbits into a gyration about a guiding centre's motion shows a bifurcation at the gravitational neutral points at the ends of the bar. Could it be that it is the phase of the gyration motion as the star enters the bifurcation region that determines which way the orbit goes? If so, we have a natural origin for the chaos that has been observed in orbits near corotation Contopoulos et al., [10]. In this paper, Sect. 2 is devoted to exact special cases in which the two dimensional motion in the galactic plane is integrable, Antonov & Shanshiev [2]. Section 3 develops the theory of guiding centre motion; when the gyration is of zero amplitude this motion is along Ergos curves which are not far from equipotential Lynden-Bell [22]. Section 4 considers the finite gyration about slightly modified Ergos curves while Sect. 5 analyses the motion near saddle points and the behaviour of the switch that directs the orbit into libration or circulation.

In the related problem in which a charged particle moves in an electromagnetic field some progress has been made in classifying the separable systems' scalar and vector potentials but even for axial symmetry Lynden-Bell [23] such classification is far from complete, although the charged Kerr Metric with $G = 0$ provides a very interesting special case Lynden-Bell [24].

## 2 Exact Special Cases

In rotating axes the equations of motion of a star in a galactic plane may be written

$$\ddot{\mathbf{R}} = \boldsymbol{\nabla}\varPhi - 2\boldsymbol{\varOmega} \times \dot{\mathbf{R}} \tag{1}$$

where $\varPhi = \psi + \frac{1}{2}\varOmega^2 R^2$ is the gravitational plus centrifugal potential measured in the sense that $\varPhi$ is large in those regions to which particles are attracted by gravitational or by centrifugal forces. Two special cases give the clue as to what to do next

1. When $\boldsymbol{\nabla}\varPhi = \mathbf{g}$ is a constant then we may orient the $y$ axis upwards, i.e., along $-\mathbf{g}$. We then have a case analogues to the $\mathbf{E} \times \mathbf{B}/B^2$ drift of plasma

physics, writing $g_y = -g = d\Phi/dy$

$$\ddot{x} = 2\Omega\dot{y} \ , \tag{2}$$

$$\ddot{y} = -g - 2\Omega\dot{x} \ . \tag{3}$$

We integrate the first and insert it into the second to find with $c$ a constant

$$\dot{x} = 2\Omega(y - c) \ , \tag{4}$$

$$\ddot{y} + 4\Omega^2(y - c + \tfrac{1}{4}g\Omega^{-2}) = 0 \ , \tag{5}$$

so $y$ oscillates harmonically about the value $c - \frac{1}{4}g\Omega^{-2} = y_0$.
In plasma physics (2) and (3) are commonly combined by writing $\zeta = x + iy$. Then

$$\ddot{\zeta} + 2i\Omega\dot{\zeta} = -ig \ ,$$

so

$$\frac{d}{dt}\left(e^{2i\Omega t}\dot{\zeta}\right) = -ige^{2i\Omega t} \ ,$$

which may readily be integrated twice to give

$$\zeta = -\tfrac{1}{2}g\Omega^{-1}t + ae^{-2i\Omega t} + \zeta_0 \ , \tag{6}$$

where $a$ and $\zeta_0$ are complex integration constants. Thus the motion consists of a circular gyration of amplitude $|a|$ and frequency $2\Omega$ about a guiding centre that moves with velocity $\mathbf{v}_d = -\frac{1}{2}g\Omega^{-1}\hat{\mathbf{x}}$ starting from point $\zeta = \zeta_0$. Notice that we may write this drift velocity in the form $\mathbf{g} \times (2\boldsymbol{\Omega})/4\Omega^2$ in analogy to $\mathbf{E}\times\mathbf{B}/B^2$ . The fact that $\mathbf{g} = \nabla\Phi$ means that the guiding centre's motion is along an equipotential but that is only true when the equipotentials are of constant curvature as we show presently. When $\mathbf{g} = \nabla\Phi$ is not constant but $\Phi$ is a non-linear function of $y$, (2) and (4) are still valid and (3) may be replaced by

$$\ddot{y} = d\Phi/dy - 4\Omega^2(y - c) = \frac{d}{dy}\left[\Phi - 2\Omega^2y^2 + 4c\Omega^2y\right] \ . \tag{7}$$

In general we now have a non-linear oscillator with an energy–like integral

$$\tfrac{1}{2}\dot{y}^2 - \Phi(y) + 2\Omega^2y^2 - 4c\Omega^2y = I = \text{constant} \ , \tag{8}$$

but let us start with the simplest case in which $g$ is expanded to first order about $y = y_0$ the trajectory of the guiding centre. Then

$$\Phi = \Phi_0 - g_0(y - y_0) + \tfrac{1}{2}\Phi_0''(y - y_0)^2 \ .$$

Equation (5) then takes the form

$$\ddot{y} + \kappa^2(y - y_0) = 0$$

where $\kappa^2 = 4\Omega^2 - \Phi_0''$ evidently

$$y - y_0 = \mathcal{I}m(ae^{i\kappa t}) \; .$$

and by (4)

$$\dot{x} = 2\Omega(y - y_0) + 2\Omega(y_0 - c) \; ,$$

$$x = \frac{2\Omega}{\kappa}\mathcal{R}e\left(ae^{i\kappa t}\right) + 2\Omega(y_0 - c)t + x_0 \; . \tag{9}$$

If we write

$$\zeta = x + i\frac{2\Omega}{\kappa}(y - y_0) \; ,$$

then

$$\zeta = (2\Omega/\kappa)ae^{i\kappa t} + v_d t + \zeta_0 \; ,$$

where the first term represents an elliptical gyration at angular frequency $\kappa$ and the remainder is the drift motion of the guiding centre at velocity

$$v_d = -g_0 2\Omega/\kappa^2$$

along $y = y_0$. In the non–linear case (8) $y$ has some mean value which we may again call $y_0$ and $\langle\dot{x}\rangle = 2\Omega(y_0 - c)$ where $\langle\dot{x}\rangle$ indicates the temporal mean.
Evidently $\dot{x} - \langle\dot{x}\rangle = 2\Omega(y - y_0)$ so $x$ executes an oscillation out of phase with $y - y_0$, making a closed curve which moves with the guiding centre. More generally again $\Phi_0''$ might depend on $x$. Then (2) and (3) would be replaced by

$$\ddot{x} = 2\Omega\dot{y} + \frac{\partial\Phi_0''}{\partial x}\tfrac{1}{2}(y - y_0)^2 \; ,$$
$$\ddot{y} = \frac{\partial\Phi_0}{\partial y} - 2\Omega\dot{x} \; .$$

if we again write $\dot{x} = 2\Omega(y - c)$ then $c$ must vary, albeit slowly, since $(y - y_0)^2$ is small. We again get

$$\ddot{y} + \kappa^2(y - y_0) = 0$$

but now $y_0$ may depend weakly on time.
We form the adiabatic invariant

$$J = \frac{1}{2\pi}\oint \dot{y}dy \; , \tag{10}$$

which will depend on $y_0$ through the value of $\kappa^2$ . We then use the invariance of $J$ and the exact conservation of the energy $\frac{1}{2}\dot{\mathbf{R}}^2 - \Phi = E_R$ to determine the small changes in $c$ and $y_0$.

2. In the above, the equipotentials were lines of constant $y$ (or almost so). More generally suppose that the equipotentials $\Phi$ = constant are curved with radius of curvature $r$ at the position considered. If $\Phi = \Phi(r)$ we may solve (1) in cylindrical polar coordinates centred at the centre of curvature. With $\phi$ the azimuthal angle we have

$$r^{-1} d/dt(r^2\dot{\phi}) = -2\Omega\dot{r}$$

so

$$r^2(\dot{\phi} + \Omega) = h = \text{constant} \ , \tag{11}$$

and

$$\ddot{r} - r\dot{\phi}^2 = \Phi'(r) + 2\Omega r\dot{\phi} \ ,$$

hence

$$\ddot{r} = d/dr\left[\Phi - \tfrac{1}{2}\left(hr^{-1} - \Omega r\right)^2\right] \ , \tag{12}$$

so

$$\tfrac{1}{2}\dot{r}^2 - \left[\Phi - \tfrac{1}{2}\left(hr^{-1} - \Omega r\right)^2\right] = E_R \ .$$

Notice that if the centre of curvature were the galaxy's centre then $r = R$. This energy is $\frac{1}{2}(\dot{r}^2 + r^2\dot{\phi}^2) - \Phi$, precisely the energy in the rotating axes. We have written motion in an axially symmetrical potential in this complicated way (in rotating axes) not merely to see the analogy with problem (1) but also because we now wish to consider problems lacking any global axial symmetry which are nevertheless steady when viewed from rotating axes. Our results are in a suitable form for applications to barred spiral galaxies and to galaxies with strong non-radial gravity fields.

We shall now generalise the above results to cases where the equipotentials are not of constant curvature but have their curvatures varying continuously along the orbits. Provided that the epicyclic motion is rapid compared with the drift motion of the guiding centre along, or almost along, the equipotential we expect an adiabatic invariant for the oscillation across the equipotentials of the form $J = (2\pi)^{-1} \oint \dot{y} dy$ . This together with the exact conservation of the energy relative to the rotating axes, gives two integrals of the motion and allows the calculation of the orbits generally and of the drift trajectories of the guiding centres in particular. In the next section we shall concentrate on finding the drift trajectories of the guiding centres. Among all possible orbits will be some for which the adiabatic invariant governing the gyration about the guiding centre is and remains zero. Thus there will be a one parameter family of non-oscillating trajectories.

## 3 Drift Trajectories – Ergos Curves

Near corotation, drift velocities are slow and guiding centre accelerations negligible, so (1) can be rewritten in the galactostrophic approximation in which Coriolis force balances the gradient of the potential

$$2\boldsymbol{\Omega} \times \dot{\mathbf{R}} = \boldsymbol{\nabla\Phi} \ , \tag{13}$$

so

$$\dot{\mathbf{R}} = \boldsymbol{\nabla}\boldsymbol{\Phi} \times \boldsymbol{\Omega}/(2\Omega^2) \ . \tag{14}$$

The drift velocity $\dot{\mathbf{R}}$ is thus along an equipotential (of constant $\Phi$) - this is just the $\mathbf{E} \times \mathbf{B}/B^2$ drift of plasma physics. However, here this approximation is unsatisfactory since it actually conflicts with the exact conservation of energy whenever $|\nabla\Phi|$ varies along an equipotential. $\dot{\mathbf{R}}^2$ as given by (14) clearly varies along an equipotential so $E_R = \frac{1}{2}\dot{\mathbf{R}}^2 - \Phi$ clearly varies along an equipotential. But $E_R$ is strictly conserved along any trajectory so the approximation that gave the drift trajectories along equipotentials conflicts with exact conservation of energy. We now give a treatment free of such conflict.

We start again but now suppose that the drift trajectories lie at small angles to the equipotentials rather than along them. Let $\hat{\mathbf{n}}(x,y)$ be the unit normal to the drift trajectories with the sense that $\hat{\mathbf{n}} \times \hat{\boldsymbol{\Omega}} \equiv \hat{\mathbf{t}}$, gives the direction of the drift velocity. $\hat{\boldsymbol{\Omega}}$ is the vector $\boldsymbol{\Omega}/\Omega$ . Then $\hat{\mathbf{n}}$ lies at a small angle to $\boldsymbol{\nabla}\boldsymbol{\Phi}$ c.f. (14). Further we shall define the curvature vector of the drift trajectories $\mathbf{K}(x,y)$ . $\mathbf{K}$ is perpendicular to the trajectory and points towards its centre of curvature from $(x,y)$ . The magnitude of $K$ is the reciprocal of the radius of curvature of the drift trajectory at $(x,y)$. A star travelling along a drift trajectory at velocity $v$ will have a transverse acceleration $\mathbf{K}v^2$ towards that centre of curvature. Taking components of (1) along the trajectory's normal $\hat{\mathbf{n}}$ we thus find

$$\mathbf{K} \cdot \hat{\mathbf{n}} v^2 = \hat{\mathbf{n}} \cdot \boldsymbol{\nabla}\Phi - 2\Omega v \ . \tag{15}$$

To simplify this notation we put $\mathbf{K} \cdot \hat{\mathbf{n}} = K$ noting that $\mathbf{K}$ and $\hat{\mathbf{n}}$ are both perpendicular to the trajectory; $K$ is either $|\mathbf{K}|$ or $-|\mathbf{K}|$ depending on the sense of the trajectory's curvature. Solving for $v$ we find

$$v = \frac{1}{K}\left[\sqrt{\mathbf{K} \cdot \boldsymbol{\nabla}\Phi + \Omega^2} - \Omega\right] = \frac{\hat{\mathbf{n}} \cdot \nabla\Phi}{\sqrt{\mathbf{K} \cdot \nabla\Phi + \Omega^2} + \Omega} \ . \tag{16}$$

Notice the close correspondence between this expression and (14) which gives $v = |\nabla\Phi|/(2\Omega)$ . Evidently if the angle between $\hat{\mathbf{n}}$ and $\boldsymbol{\nabla}\Phi$ is small enough to have its square neglected, and if $|\mathbf{K}\nabla\Phi| \ll \Omega^2$ the two expressions become equal. However (16) is exact while (14) was approximate. We now use exact energy conservation relative to the rotating axes

$$E_R = \tfrac{1}{2}v^2 - \Phi = \tfrac{1}{2}\frac{(\hat{\mathbf{n}} \cdot \boldsymbol{\nabla}\Phi)^2}{\left[\sqrt{\mathbf{K} \cdot \boldsymbol{\nabla}\Phi + \Omega^2} + \Omega\right]^2} \ - \Phi \equiv \mathcal{E}(x,y) \ . \tag{17}$$

The function $\mathcal{E}(x,y)$ is called the Ergos (Lynden-Bell, [22]). The definition is implicit since $\hat{\mathbf{n}}$ is the normal to the trajectories along which $\mathcal{E}$ is constant and whose curvatures are given by $\mathbf{K}(x,y)$ . So far all is exact; the Ergos curves along which $\mathcal{E}$ is constant give the drift trajectories of the (zero amplitude) guiding centres. Now for any function $F$ that is $-\Phi$ or any better approximation to the Ergos, writing suffixes to denote differentiation, and

$$s = (F_x^2 + F_y^2)^{\frac{1}{2}} \ , \tag{18}$$

$$\hat{\mathbf{n}} = -s^{-1}(F_x,\ F_y);\ \hat{\mathbf{n}} \times \hat{\boldsymbol{\Omega}} = \hat{\mathbf{t}}\ , \tag{19}$$

we have

$$\mathbf{K} = \left(\hat{\mathbf{t}} \cdot \boldsymbol{\nabla}\right) \hat{\mathbf{t}} = s^{-3} \left(F_x^2 F_{yy} - 2F_x F_y F_{xy} + F_y^2 F_{xx}\right) \hat{\mathbf{n}}\ . \tag{20}$$

Wherever $|\mathbf{K} \cdot \boldsymbol{\nabla}\Phi|$ is not as large as $\Omega^2$ it is easy to find approximations to the Ergos. At zero order we use $-\Phi$ for $F$ and calculate first approximates to $\hat{\mathbf{n}}$ and $\mathbf{K}$ from the above formulae. Substituting them into (17) we find a first approximation to the Ergos $\mathcal{E}_1(x, y)$ . Using $\mathcal{E}_1$ for $F$ in the above formulae we calculate 2nd approximations to $\hat{\mathbf{n}}$ and $\mathbf{K}$ and putting them into (17) we get $\mathcal{E}_2(x, y)$ . Near corotation this will converge quite quickly to give the Ergos and the level surfaces of it give the Ergos curves along which the guiding centre trajectories lie. For related work on such systems see Antonov & Shanshiev [2]. Very close to gravitational neutral points where $\nabla\Phi = 0$ it is easiest to calculate the Ergos curves as trajectories with zero gyration directly. Figs. 1 and 2 give the equipotentials and the Ergos Curves.

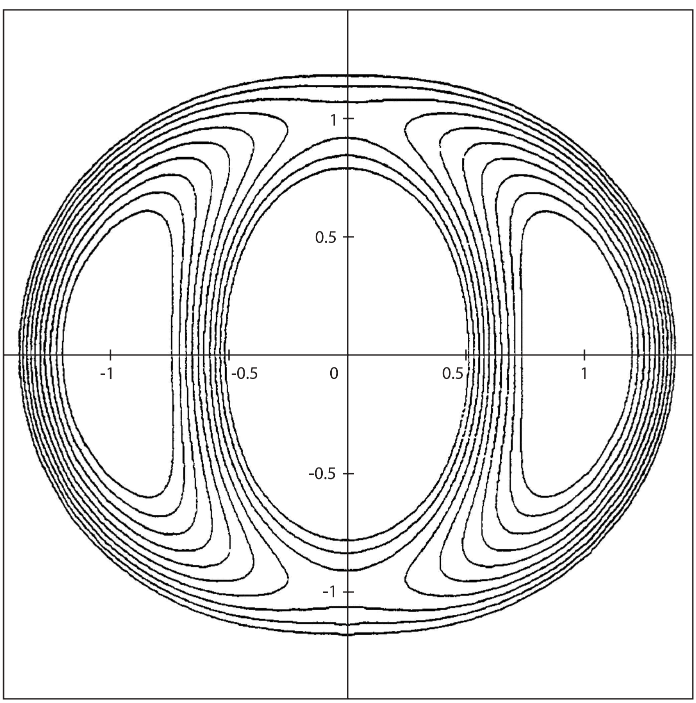



**Fig. 1.** Equipotentials of $\Phi = \psi + \frac{1}{2}\Omega^2 R^2$ where $R^2 = x^2 + y^2$; $\psi = \mathcal{G}M(b + s)^{-1}\left[1 - 0.02b^2(x^2 - y^2)s^{-4}\right]$ ; and $s^2 = R^2 + b^2$. The angular velocity of the bar, $\Omega$, is chosen so that $\Omega^2 s = \mathcal{G}M/(b + s)^2$, that is $\Omega^2 = \mathcal{G}Mb^{-3}/(4 + 3\sqrt{2})$. In the diagram $\mathcal{G}M = 1$ and $b = 1$.

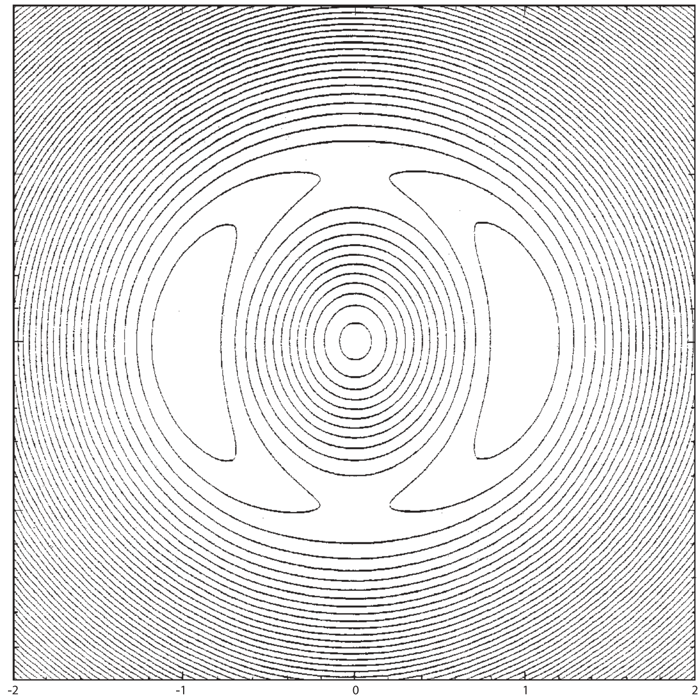



**Fig. 2.** Ergos curves for zero gyration and the potential of Fig. 1.

# 4 The Gyration Adiabatic Invariant

At any point $\mathbf{R}_0$ the equipotentials have some curvature $K_\circ$ and near there $\Phi$ can be approximated as being a function of $r$ the distance to the centre of curvature. Thus, in a region near $\mathbf{R}_0$ the angular momentum about that centre of curvature will be approximately conserved. Taking the cross product of (1) by $\mathbf{r}$ the vectorial distance from that centre of curvature and using $\dot{R} = \dot{r}$ we find

$$d/dt(\mathbf{r} \times \dot{\mathbf{r}}) = -2\mathbf{r} \times (\boldsymbol{\Omega} \times \dot{\mathbf{r}}) + O(\epsilon^2) \ ,$$

so

$$\mathbf{r} \times (\dot{\mathbf{r}} + \boldsymbol{\Omega} \times \mathbf{r}) = \mathbf{h} + O(\epsilon^2) \ , \tag{21}$$

as in (11), but now $h$ is only approximately constant locally. (21) will be just as true of the motion of the guiding centre as it is of the motion of the star that gyrates about that centre. Let the guiding centre be at $r_0$ and the star at $r_0 + \eta$ then working to first order in $\eta$ writing $r = r_0 + \eta$ in (12)

$$\ddot{\eta} + \kappa^2 \eta = -(\hbar r_0^{-2} - \Omega)\delta h \ ,$$

where $\kappa^2 = -d^2\Phi/dr^2 - h^2 r_0^{-3} + \Omega^2 r_0$ . We could have chosen to compare the motion of our star with that of a guiding centre with the same $h$ and put

$\delta h = 0$ but as neither $h$ nor $\delta h$ are quite constant we have chosen not to do that. Evidently $\eta$ vibrates harmonically about $\kappa^{-2}(\Omega - hr_0^{-2})\delta h = \eta_0$ . We shall now assume that this $\eta$ vibration is sufficiently fast that the corresponding action is adiabatically invariant. So the invariant is

$$J = \frac{1}{2\pi}\oint \dot{\eta} d\eta = \oint \sqrt{2[E_R + \Phi] - (hr^{-1} - \Omega r)^2} dr$$
$$= \Delta E_R/\kappa$$

where $\Delta E_R$ is the excess energy above that of the guiding centre. The integral is evaluated with $E_R$ and $h$ fixed and with $\Phi = \Phi(r, \phi, \mathbf{R}_0)$ only weakly dependent on $\phi$ and expanded about the point $R_0$ and $\phi$ . In the integration $R_0$ and $\phi$ are held fixed and only $r$ varies. Henceforth any dependence on $\phi$ may be incorporated into the $R_0$ dependence. Thus we find

$$J = J(E_R, h, \mathbf{R}_0) \ .$$

$\Delta E_R$ and $J$ are second order in the displacement from the guiding centre. We are now able to give a correction of this order to the guiding centre's motion which we earlier determined in the limit when $J$ was zero. When $J$ is non-zero the vibration about the guiding centre has extra energy $\Delta E_R = \kappa J$. While $J$ is fixed; $\kappa$ still varies from point to point. Thus the effective potential for the guiding centres motion is

$$\widetilde{\Phi} = \Phi - \kappa J \ ,$$

so that the energy of the total motion is

$$E_R = \tfrac{1}{2}\dot{\mathbf{R}}^2 - \Phi = \tfrac{1}{2}\dot{\mathbf{R}}_0^2 - \widetilde{\Phi} \ ,$$

where $\dot{\mathbf{R}}_0$ is the motion of the guiding centre. Thus the Ergos curves for guiding centres of given $J$ should be calculated with $\widetilde{\Phi}$ replacing $\Phi$ . Fig. 3 shows a banana orbit in which one can see the gyration especially near the ends of the banana. Fig. 4 shows an orbit that starts librating in a banana close to the critical ergos curve but then switches to circulation outside corotation.

## 5 Is the Saddle-Point Switch Chaotic?

When in the 1960's Michel Hénon [19], [18] and George Contopoulos [8] discovered the fascination of the onset of chaos in stellar dynamical orbits I saw that a new branch of mathematics would develop, (Drazin [14]), but by that time I was more interested in the astrophysical problems cast up by astronomy than in the purely mathematical ones. I have never regretted that decision, though I have watched with admiration the developments pioneered by my more mathematical colleagues. One of the early examples of chaos was in Doug Allen's thesis [1] on the behaviour of coupled disk dynamos. The problem was suggested by Bullard and its solutions gave some indication of why the Earth's magnetic field suffers chaotic reversals. Later I learned of the pioneering studies of Mary

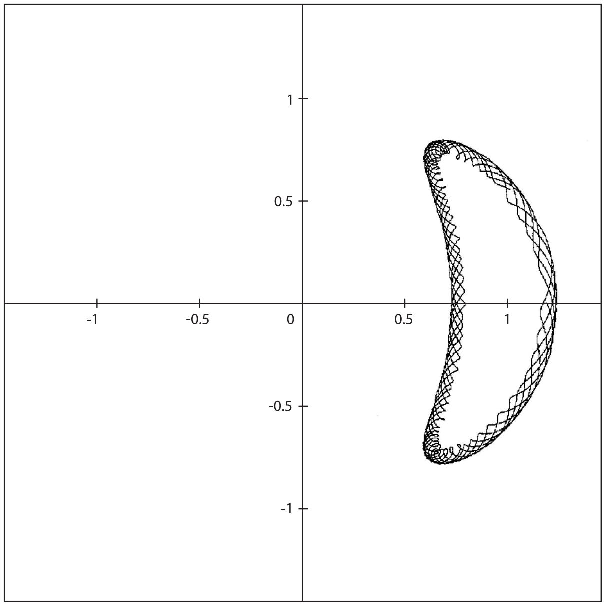



**Fig. 3.** A banana orbit showing the effects of gyration about the moving guiding centre especially near the ends of the banana.

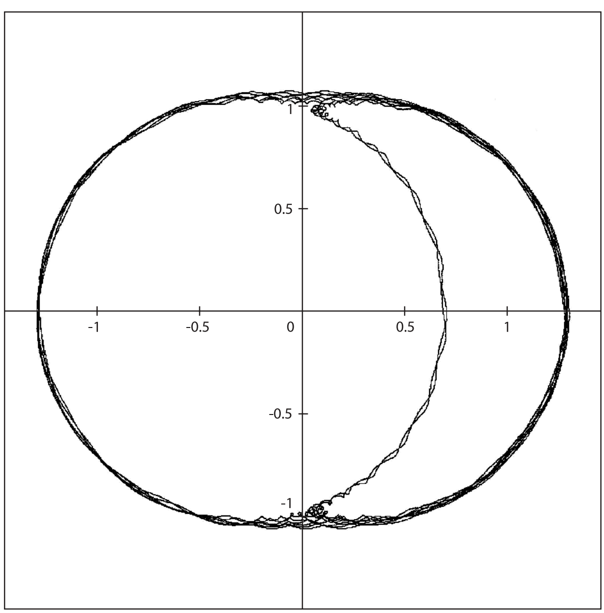



**Fig. 4.** A banana orbit very close to the critical Ergos curve, which switched from a librating orbit to one circulating outside corotation. Although integration was continued much longer it did not switch back. Either the orbit must hit a very small hole to cross back or the inaccuracy of the integrator allowed a small change in the guiding–centre motion so that it no longer came close to the critical switch.

Cartwright [5] who came upon the chaotic behaviour in the more mathematical context of differential equations. A deep mathematical study of the conditions that generate chaos in such systems was made by Colin Sparrow [27] under the aegis of Peter Swinnerton-Dyer and DLB gained a taste for their mathematical rigour by attending some of their lectures on chaos. What little he remembers involved orbits that continually came back into a critical region from which they could emerge in one of several different directions. It was the critical switching between these that led to chaos in the solutions. Over the years he has heard George Contopoulos talk about chaos many times, and chaos near corotation in orbits that get close to the ends of the bar in barred spiral galaxies has often been found. When George spoke on the subject at the Saltsjobaden Meeting in December 1995 [10], DLB had the belief that the saddle points in the gravitational potential provided just that critical switch with two very different outcomes that Sparrow needed. I thought the gyrations of the stellar orbit about its guiding centre would provide just that wobble between one side of the separatrix and the other needed to give chaos. The pressure of preparing this talk provided the stimulus needed to work this out properly. We start by analysing the switch at one of the gravitational points shown in Fig. 1. Centering our coordinates $x, y$ on the upper saddle point $\Phi$ may be expanded for $x, y$ small in the form

$$\Phi = \Phi_0 + \tfrac{1}{2}\alpha^2 x^2 + \tfrac{1}{2}\beta^2 y^2 \ ,$$

so the equations of motion (1) take the form

$$\ddot{x} = -\alpha^2 x + 2\Omega\dot{y} \ ,$$

$$\ddot{y} = \beta^2 y - 2\Omega\dot{x}$$

writing $D$ for $d/dt$ we see that

$$(D^4 + \omega_0^2 D^2 - \alpha^2\beta^2)x = 0 \ ,$$

where $\omega_0^2 = \alpha^2 + 4\Omega^2 - \beta^2$, and $y$ obeys the same equation. In practice $\alpha^2 + 4\Omega^2 - \beta^2 > 0$. Writing $D = iw$ we see that, for $w^2$ there is one positive root

$$w^2 = w_1^2 = \tfrac{1}{2} w_0^2 \left(1 + \sqrt{1 + 4\alpha^2\beta^2\omega_0^{-4}}\ \right)$$

and a negative one with

$$-\omega^2 = \gamma^2 = 2\omega_0^{-2}\alpha^2\beta^2 / \left(1 + \sqrt{1 + 4\alpha^2\beta^2\omega_0^{-4}}\ \right) \ .$$

The dying solution, $\gamma > 0$, $e^{-\gamma t}$ corresponds to a contraction of the points along the separatrix line from upper left or bottom right while the growing $e^{\gamma t}$ solution corresponds to expansion along the separatrix line from the saddle both to lower left and upper right. Together these motions give $x = \gamma(Ae^{-\gamma t} + Be^{\gamma t})$, $2\Omega y = (\gamma^2 - \alpha^2)Ae^{-\gamma t} + (\gamma^2 + \alpha^2)Be^{\gamma t}$ where $A$ & $B$ are arbitrary constants with the separatrices given by $B = 0$ and $A = 0$ respectively. This flow is drawn in Fig. 6. At the saddle the flow switches to left or to right depending on the sign of $B$ which decides on which side of the separatrix the guiding centre approaches. However, superposed on these motions is an elliptical gyration due to the real roots $\omega^2 = \omega_1^2$, these give $x = \omega_1 \mathcal{C} e^{i\omega_1 t}$ and $2\Omega y = (\omega_1^2 - \alpha^2) i \mathcal{C} e^{i\omega_1 t}$ where $\mathcal{C}$ is an arbitrary complex constant and the real $x$ and real $y$ are the real parts of the

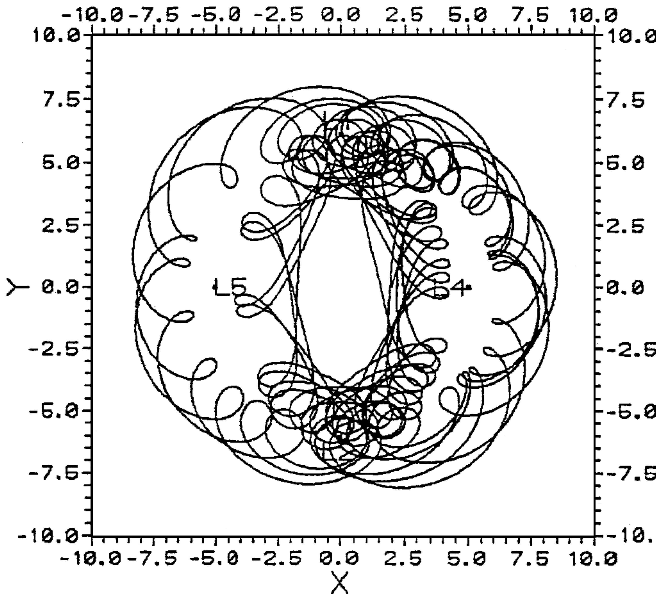



**Fig. 5.** A chaotically switching orbit of large gyration amplitude computed by Contopoulos et al 1996.

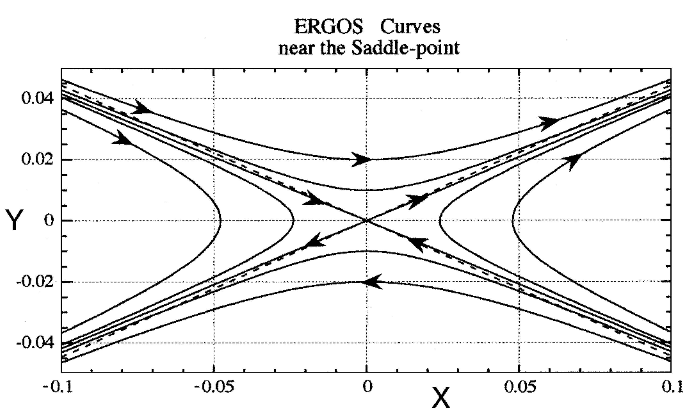



**Fig. 6.** Guiding centre flow close to the saddle point. Critical equipotentials (*shown dashed*) are close to the critical Ergos curves.

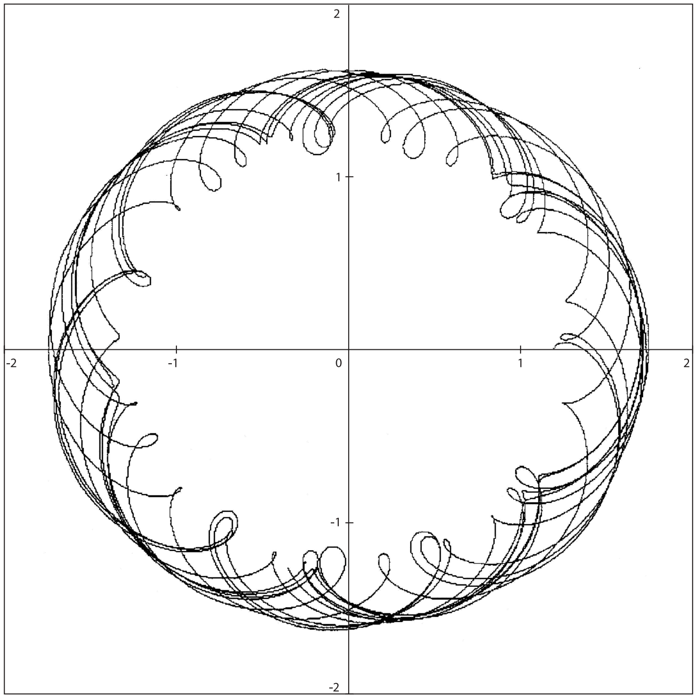



**Fig. 7.** An orbit outside corotation shows gyrations of slightly variable amplitude.

expressions given. Interestingly this elliptical[1] gyration continues unaffected by the saddle point.

Thus the switch to left or right is determined **not** by the position of the star but by the position of its guiding centre. DL-B's concept at the start of this investigation was that the switch would act on the star's position, so that the phase of the gyration as the star approached the saddle point would be crucial. Now this concept is seen to be false there is no random switching because the guiding centres follow the ergos curves. What then is the origin of the apparent switching of orbits seen in Figs. 4 & 5?

Three possibilities are

1. At finite gyration amplitudes there are resonances between the gyration and the motions of the guiding centres which lead to oscillations in the value of $J$ and of the energy of the guiding centre's orbit which allow it to cross the separatrix before approaching the saddle–point switch.
2. The zero gyration motion of the guiding centre along an ergos curve is itself unstable.
3. For bars with significant non-radial forces the motions along the ergos curves are too rapid for the good conservation of the adiabatic invariant. Accurate separation between a guiding centre motion and a gyration is not possible

[1] For Fig. 6 the ellipse is almost round being only 1% flattened in $y$.

except close to the saddle–points. Orbits close to the separatrix will return on different sides of it on different approaches to the saddle–points.

Thus we have been unable to **isolate** the origin of the apparent randomness in the switching. However, we hope we have added some understanding of the orbits and of their integrals of motion.

Figure 7 shows an orbit that circulates "backward" outside corotation. Notice that even at the same azimuth there are small differences in the gyration amplitude; this may be due to inexact conservation of the adiabatic invariant.

## References


1. D.W. Allen: Proc. Camb. Phil. Soc. 68, 671 (1962)
2. V. A. Antonov, & F.T. Shanshiev: 1993 Celestial Mechanics & Dynamical Astronomy, **59**, 209 (1994)
3. R.B. Berman & J.W. Mark: ApJ, **216**, 257 (1977)
4. J.J. Binney & S.D. Tremaine: *Galactic Dynamics* (Princeton University Press, Princeton 1987)
5. M. Cartwright & D.E. Littlewood: J.London Math.Soc., 20, 180, (1946)
6. S. Chandrasekhar: Principles of Stellar Dynamics p132, Dover NY 196 (1942)
7. G.L. Clarke: MNRAS **97**, 182 (1937)
8. G. Contopoulos: Astron. J., **76**, 147 (1971)
9. G. Contopoulos & P.O. Vandervoort: ApJ, **389**, 119 (1992)
10. G. Contopoulos, N. Voglis & C. Efthymiopoulos: *B*arred Galaxies & Circumnuclear Activity eds. Aa Sandqvist & P.O. Lindblad, Proceedings of Nobel Symposium 98, Springer, Berlin (1996)
11. P.T. de Zeeuw & D. Merritt: ApJ, **267**, 571 (1983)
12. P.T. de Zeeuw: MNRAS **216**, 273 (1985)
13. P.T. de Zeeuw: MNRAS **216**, 599 (1985)
14. P.G. Drazin: *Non Linear Systems* (Cambridge University Press, Cambridge 1992)
15. A.S. Eddington: MNRAS **76**, 37 (1915)
16. K.C. Freeman: MNRAS, **134**, 1 (1966)
17. L.S. Hall: Physica 8D, 90 (1983)
18. M. Hénon: Quart. App. Math, **27**, 291 (1969)
19. M. Hénon & C Heiles: Astron J, **69**, 73 (1964)
20. G.G. Kuzmin: Astron. Zh **33**, 27 (1956)
21. D. Lynden-Bell: MNRAS **124**, 95 (1962)
22. D. Lynden-Bell: *Barred Galaxies & Circumnuclear Activity* eds. Aa Sandqvist & P.O. Lindblad, Proceedings of Nobel Symposium 98, (Springer, Berlin, 1996)
23. D. Lynden-Bell: MNRAS, **312**, 301 (1999)
24. D. Lynden-Bell: *New Developments in Astrophysical Fluid Dynamics*, (eds. Dalsgaard & Thompson), C.U.P., Cambridge, in press, (2002), (astro-ph/0207064)
25. D. Lynden-Bell: MNRAS, accepted, (2003), (astro-ph/0210417)
26. I. Marshall & S. Wojciechowski: J.Math.Phys., **29**, 1338 (1988)
27. C. Sparrow: *The Lorentz Equations Bifurcations Chaos and Strange Attractions*, (Springer Verlag, New York App Math, 41, 1982)
28. P.O. Vandervoort: ApJ, **232**, 91 (1979)
29. E.T. Whittaker: *Analytical Dynamics*, (Cambridge University Press, Cambridge, 1959)

# Stellar Dynamics and Molecular Dynamics: Possible Analogies

Andrea Carati and Luigi Galgani

Università di Milano, Dipartimento di Matematica, Via Saldini 50,
I–20133 Milano, Italy

**Abstract.** In stellar dynamics one is accostumed to deal with the Lynden–Bell distribution, which presents two peculiar characteristics: a) it resembles the quantum Fermi–Dirac distribution, and b) describes a state of metaequilibrium, that is expected to evolve on much longer time scales to a standard Maxwell–Boltzmann equilibrium. Here it is illustrated how an analogous situation seems to occur in molecular dynamics, described within the context of classical mechanics. The problem concerns the contribution of the internal degrees of freedom, typically the vibrations, to the specific heat; correspondingly, the metaequilibrium state leads to a Bose–Einstein–like rather than to a Fermi–Dirac–like distribution. We also point out that in molecular dynamics the "nonclassical" features seem to be related to the fact that the evolution of the energy is dominated by the presence of rare but conspicuous jumps, as in processes of Lévy type; this too has some analogies with stellar dynamics.

## 1 Introduction

About thirty years ago it occurred to one of the present authors, in collaboration with A. Scotti and C. Cercignani [1][2], to observe a quantum–like feature in the problem of the specific heats studied in the context of classical mechanics. The model considered was that of Fermi–Pasta–Ulam [3], describing a one–dimensional crystal with nonlinear interactions between adjacent atoms. The observation was that, if the specific energy was small enough and the energy was given initially to the lowest–frequency mode, the distribution of energy among the modes, estimated by numerical solutions of the equations of motion, turned out to have a Planck–like form. Moreover, to a great astonishment, even the action entering such a distribution turned out to be of the order of magnitude of Planck's constant. It took some months to become convinced that the latter fact was neither a mistake nor an accident, as we briefly recall now. The relevant point is that, in the model, realistic molecular parameters had been introduced; for example, for a model of crystal Argon, one introduces the mass $m$ of Argon and a realistic Lennard–Jones interatomic potential $V(r) = 4\epsilon[(\sigma/r)^{12} - (\sigma/r)^6]$, with certain parameters $\epsilon$ and $\sigma$ given in the literature. Now, one immediately checks that the action $A$ naturally built up from the given parameters $m, \epsilon, \sigma$ is just $A = \sqrt{m\epsilon}\,\sigma$; on the other hand it turns out that for realistic parameters there exists the relation $A = 2Z\hbar$ where $\hbar$ is the (rationalized) Planck's constant and $Z$ the atomic number. This remark explains how in classical models of molecular dynamics Planck's constant is introduced, so to say by hands, through the values of the molecular parameters.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 44–55, 2003.

In such a way an adventure was started, which consisted in trying to prove that, at least in the problem of the specific heats (see however [4], where the electrodynamics of point particles is also discussed), classical mechanics might not be inconsistent with quantum mechanics, or at least in understanding how something like this might make sense. Having explained in which way Planck's constant had been introduced, somehow by hands, through the molecular parameters, the main problem was then to understand how the dynamics itself, described by Newton's equations with given potentials, might make the job. This is the reason why intense contacts were started with scientists working in the field of dynamical systems, from mathematics to celestial mechanics. Perhaps the first among them was G. Contopoulos, with whom an everlasting friendship and scientific collaboration was initiated. The adventure had since then several phases, with several incursions into purely mathematical aspects of the theory of dynamical systems; a long collaboration with G. Benettin and A. Giorgilli took place, and finally the first of the present authors joined the party.

The present phase seems to be characterized by the realization that there should exist two (or several) well distinct relaxation times in the problem, as is now familiar in the physics of glasses or spin glasses. The point we want to stress here is that the existence of two relaxation times is a familiar feature in stellar dynamics too. Indeed, one there refers to a first "violent" relaxation to a metaequilibrium state of the type of Lynden–Bell [5] (during which the collisions can be neglected), that should then be followed by an extremely slow relaxation to a standard equilibrium state governed by the collisions. Moreover, such a phenomenon of the existence of a metaequilibrium state turns out to occur, in stellar dynamics, just in conjunction with the appearence of a quantum–like feature, namely the Fermi–like distribution of Lynden–Bell. These two features, namely the existence of a metaequilibrium state and its actual quantum–like aspect, constitute the analogies of stellar dynamics with the problem of the specific heats in classical mechanics which motivated the present talk, mainly addressed to people working in the field of stellar dynamics.

## 2 Planck's Law, Its Interpretation by Einstein, and the Points of View of Boltzmann and of Nernst

Planck's law is concerned with the mean energy $U$ of a system of $N$ harmonic oscillators of angular frequency $\omega$ at absolute temperature $T$. In terms of inverse temperature $\beta = 1/k_B T$ and of the quantum of energy $\hbar\omega$, where $k_B$ is the Boltzmann's constant, it asserts that the mean energy $U$ has the form

$$U = N \frac{\hbar\omega}{e^{\beta\hbar\omega} - 1} \ . \tag{1}$$

The relevant feature of this law is that in the limit of high temperatures or low frequencies, i.e. for $\beta\hbar\omega \ll 1$, it leads to the "classical" value $U = N k_B T$, independent of frequency (this realizes the so–called equipartition of energy), while it gives degeneration, i.e. the vanishing of $U$ (in a frequency–dependent way), for $\beta\hbar\omega \gg 1$.

Everyone knows how Planck's law is usually deduced, by the standard equilibrium argument using the Maxwell–Boltzmann principle, if energy is assumed to be quantized, i.e. if one admits that the allowed values for the energy $E$ are $E_n = n\hbar\omega$, $n = 1, 2, \cdots$. This standard argument in fact constitutes the second deduction given by Planck himself; a variant of it, which takes into account the zero–point energy $N\hbar\omega/2$, comes about if the energy levels are assumed to be given by $E_n = (n + 1/2)\hbar\omega$.

So, Planck's law turns out to be a consequence of quantization. Conversely, it was shown by Poincaré [6] (in one of his last papers, that he wrote under the stimulus of the discussions at the first Solvay Conference [7]) that quantization is necessary if one has to recover Planck's law at all. Actually, an accurate analysis of the paper of Poincaré shows that the situation is not so clear, and for example Einstein never was convinced of this necessity of quantization. This is witnessed by some remarks he made in his scientific autobiography [8]. Indeed, after having recalled how he himself had "*showed in a definitive and direct way that it is necessary to attribute a certain immediate concreteness to Planck's quanta and that, under the energetic aspect, radiation possesses a sort of molecular structure*", after a few lines he adds: "*This interpretation, that almost all contemporary physicists consider as essentially definitive, to me appears instead as a simple provisional way out*".

What did Einstein have in mind in saying these words? In our opinion the answer is found in his contribution to the Solvay Conference, where he showed how Planck's law can be obtained by arguments which make no reference to quantization at all. The key point is a physical interpretation of the procedure that had been followed by Planck in the first deduction of his law, on October 19, 1900. Let us recall that in that paper Planck had obtained his formula as a solution of the ordinary differential equation (we are using here a contamination of the notations of Planck and of Einstein)

$$\frac{\mathrm{d}U}{\mathrm{d}\beta} = -\big(\hbar\omega\, U + \frac{U^2}{N}\big)\,, \tag{2}$$

to which he had arrived, with no real physical interpretation, by a purely formal interpolation between two limit equations, well adapted to the cases of high frequencies and low frequencies respectively. What Einstein did, was to split such an equation into a system of two equations, namely

$$\frac{\mathrm{d}U}{\mathrm{d}\beta} = -\sigma_E^2 \tag{3a}$$

$$\sigma_E^2 = \hbar\omega\, U + U^2/N\,, \tag{3b}$$

where there appears a further quantity $\sigma_E^2$ having a well definite physical meaning, namely the variance of energy. Indeed the former equation (26a), relating specific heat to variance of energy, had been discovered in the year 1903 by Einstein himself in one of his first papers, as an identity in the canonical ensemble, and was conceived by him as a kind of a general thermodynamic relation that should have some more general validity. In his mind, the second relation (77) should instead have a dynamical character, and might in principle be deducible from a microscopic dynamics. In his very words [7]: these two relations "*exhaust the thermodynamic content of Planck's*" formula; and "*a mechanics compatible with the energy fluctuation* $\sigma_E^2 = \hbar\omega U + U^2/N$ *must then necessarily lead to Planck's*" formula. So the main idea is that the energy exchanged with a reservoir, i.e. the specific heat $\frac{\mathrm{d}U}{\mathrm{d}\beta}$, should be related to the energy fluctuations. In turn, the functional dependence of the energy fluctuations on the mean energy should fix the functional form of the mean energy in terms of temperature and of frequency. Clearly, no reference to quantization is made here.

Another key point enters now, which goes back to Boltzmann; we refer to the role of nonequilibrium. Indeed, being confronted with the phenomenological lack of equipartition of energy in crystals and in polyatomic molecules, Boltzmann conceived the idea that in such cases one was actually dealing with situations in which equilibium had not yet been reached. In his very words: [9] "*The constituents of the molecule are by no means connected together as absolutely undeformable bodies, but rather this connection is so intimate that during the time of observation these constituents do not move noticeably with respect to each other, and later on their thermal equilibrium with the progressive motion is established so slowly that this process is not accessible to observation*". In modern terms, the situations of nonequipartition of energy should be understood as analogous to the metaequilibrium situations occurring in glasses and spin glasses.

Let us recall that the idea of Boltzmann, according to which the situations of nonequipartition would correspond to states of metaequilibrium, was pursued for several years by Jeans [10], with the explicit aim of avoiding a recourse to quantization. But the work of Poincaré on the necessity of quantization made so strong an impression on him that he found himself forced to make a public retractation [11] and to abandon any further attempt in that direction. We will recall below how the problem was reopened much later, in the year 1954, by the work of Fermi, Pasta and Ulam. We would also like to mention that the idea of Boltzmann that there should exist a "time–dependent specific heat" is today accepted as a trivial fact, with no mention to Boltzmann at all (see for example [12]).

We add now here some further comments. The first remark is that there seems to be a strong relation between the point of view of Boltzmann and that of Einstein: the key point is the role of the dynamics in the problem of the specific heats. According to Boltzmann, what is relevant for the specific heat of a body is not the energy it possesses, but rather the energy it can exchange through the dynamical interaction with a heat reservoir within a given observation time. In turn, the latter energy, the exchangeable one, is related to the dynamical fluctuations of the energy of the body. This is in fact essentially the statement of the well known Fluctuation Dissipation Relation (see for example [13], of which the relation (26a) of Einstein seems to be a precursor. In fact, the Fluctuation Dissipation Relation has a form very similar to (26a), the main difference being that it involves quantities having a dynamical character. Consider a system at inverse temperature $\beta$, and denote by $E(t)$ its energy at time $t$, and by $U(t)$ the mean energy exchanged up to time $t$ with a reservoir at inverse temperature $\beta + \mathrm{d}\beta$. Then, the Fluctuation Dissipation Relation reads

$$\frac{\mathrm{d}U}{\mathrm{d}\beta} = -\frac{1}{2}\langle\big(E(t) - E(0)\big)^2\rangle \ , \tag{4}$$

the mean $\langle\cdot\rangle$ being taken with respect to initial data with a Maxwell–Boltzmann distribution at inverse temperature $\beta$. From this, the static relation (26a) of Einstein is recovered at times so large that the autocorrelation of the energy vanishes, so that $E(t)$ and $E(0)$ become independent and thus the quantity $1/2\langle\big(E(t) - E(0)\big)^2\rangle$ reduces to the static canonical variance $\sigma_E^2$.

A final comment concerns the role of the Maxwell–Boltzmann distribution as a statistical measure for the initial data, irrespective of the dynamics. The fact is that, for a system of independent harmonic oscillators distributed according to Maxwell–Boltzmann, one has for the mean energy $U$ the value $Nk_BT$, i.e. formally equipartition of energy. But such an equipartition, as Boltzmann would say, concerns the mechanical energy possessed by the system just in virtue of the choice of the initial data, and has a priori nothing to do with the thermodynamic energy, which should be defined as the exchangeable one within the observation time. The latter, i.e. the thermodynamic or exchangeable energy, is instead measured by the dynamical fluctuations of energy. The fact that the initial distribution of energy presents a nonvanishing variance $\sigma_E^2$ is of no relevance for the specific heat, which depends on the exchangeable energy, i.e. on the dynamics.

This remark explains how one can have a situation in which there is both equipartition of energy in relation to the initial data, and Planck's law in relation to the exchangeable energy, as was first conceived by Nernst [14] in an extremely deep, almost unknown, work (see also [15]). In particular, Nernst also introduced a deep conception of the energy which, in the sense of Boltzmann, does not contribute to the specific heat: on the one hand it should be characterized as being, from the dynamical point of view, of ordered type (*geordnete*); on the other hand it would constitute a classical analog of the quantum zero-point energy. It is worth recalling the argument by Nernst. He assumes that the quantum of energy $\hbar\omega$ plays the dynamical role of a stochasticity threshold for the harmonic oscillators; the motions would be of ordered type below threshold and of disordered type (*ungeordnete*) above it. Furthermore, he assumes that the oscillators are distributed according to Maxwell–Boltzmann, so that one has equipartition for their mechanical energy. Now, one immediately computes the mean disordered energy (per oscillator) $E_1$, namely the mean energy conditioned by $E > \hbar\omega$ and the mean ordered energy (per oscillator) $E_0$, namely the mean energy conditioned by $E < \hbar\omega$, and one finds

$$E_1 = k_BT + \hbar\omega \quad , \quad E_0 = k_BT - \frac{\hbar\omega}{e^{\hbar\omega/k_BT} - 1} \quad . \tag{5}$$

Similarly one finds that the fraction $n_1$ of oscillators above threshold is $n_1 = \exp(-\hbar\omega/k_B T)$. Then the exchangeable energy $U$ can be assumed to be defined by $U = N n_1 (E_1 - E_0)$, which coincides with Planck's law.

## 3 The Fermi–Pasta–Ulam Problem

The problem of a dynamical foundation for the principle of equipartition of energy in classical mechanics was reopened in the year 1954 by Fermi, Pasta and Ulam [3]. The interest of Fermi for this problem was indeed a rather old one, since it goes back to his work [16] of the year 1923 (which is sometimes misunderstood in the literature; see however [17]), where he improved the theorem of Poincaré on the integrals of motion of a Hamiltonian system. So Fermi came back to the problem when for the first time he had the facilities of a computer for the numerical integration of the equations of motion of a rather large system of particles.

As mentioned above, Fermi, Pasta and Ulam considered a system of $N$ points (atoms) on a line, with a nonlinear interaction between adjacent atoms and certain boundary conditions; typically, the positions of the extreme atoms were kept fixed and the number $N$ of moving atoms was 64. The interaction potential energy had the form $V(r) = r^2/2 + \alpha r^3/3 + \beta r^4/4$, with given constants $\alpha, \beta$. For $\alpha = \beta = 0$ one has a linear system which, by a familiar argument, is equivalent to a system of $N$ uncoupled harmonic oscillators (normal modes) having certain frequencies. The problem is then how many normal modes take part in the energy sharing, which should occur in virtue of the nonlinear interaction. The authors considered initial conditions in which the energy was given just to the lowest normal mode (i.e. to the mode of lowest frequency), and the aim was to observe, by numerical solutions of the equations of motion, the rate of the flow of energy towards the modes of higher frequency, which was expected to occur in order to establish the equipartition of energy among all the modes. They found the unexpected result that, up to the times considered, the energy appeared to be distributed just among a packet of normal modes of low frequency without flowing to the high frequency modes. They also gave a figure reporting the mean (in time average) energy versus frequency, which exhibited an exponential decay.

After this original work two more works had, in our opinion, a particularly relevant role, namely that of Izrailev and Chirikov [18] of the year 1966, and that of Bocchieri, Scotti and Loinger [19] of the year 1970. F.M. Izrailev and B. Chirikov understood that there existed the problem of an energy threshold. By analogy with the situations occurring in perturbation theory, in connection with the existence of ordered motions in the sense of Kolmogorov, Arnold and Moser, they conjectured that equipartition would be obtained if the initial energy $E$ was larger than a certain threshold energy $E^c$. Thus the result of Fermi, Pasta and Ulam was explained as being due to the fact that only small energies, with $E < E^c$, had been considered. The crucial point is then to understand how does the critical energy $E^c$ depend on the number $N$ of degrees of freedom, because in situations of physical interest $N$ should be of the order of the Avogadro number. So one has to look at the specific critical energy $\epsilon^c = E^c/N$ (we are using here the symbol $\epsilon$ for the specific energy, with no relation to the parameter $\epsilon$ of the Lennard–Jones potential mentioned in the Introduction). The authors had clearly in mind the idea that one might prove that $\epsilon^c \to 0$ as $N \to \infty$ (at least for initial excitations of the high frequency modes). Indeed, if this were true, then in any physically meaningful system one would always have equipartition of energy. Then everybody would be happy, because this would prove that classical mechanics predicts a wrong result, as everyone has learned at school.

P. Bocchieri, A. Scotti and A. Loinger (working with a Lennard–Jones interaction potential) gave a strong indication in the sense that, on the contrary, the specific energy threshold $\epsilon^c$ should tend to a finite nonvanishing value in the limit $N \to \infty$. Shortly after such a work, in the paper [1] it was shown that, just in situations in which according to Bocchieri, Scotti and Loinger one does not have equipartition, the distribution of energy among the normal modes has a Planck-like form, and even with an action of the order of magnitude of Planck's constant. Thus the adventure was started of looking for a deeper understanding of the relations between classical and quantum mechanics.

After such "old" works, many other works followed (see for example [20] and [21]), mostly with the intent of establishing whether $\epsilon^c \to 0$ as $N \to \infty$, or not. The theoretical framework also changed a lot. Indeed, initially reference was made to KAM theory, while later the point of view of N.N. Nekhoroshev entered the game [22]. The attention was thus shifted towards the idea that one might always have equipartition as $t \to \infty$, but with the possibility that the relaxation time might increase exponentially fast as the specific energy decreases.

From this point of view, a recent relevant result was given, in our opinion, in the paper [23]. Here it is confirmed that the results depend on the specific energy $\epsilon$, and actually in the following way. There exists a specific energy threshold $\epsilon^c$ such that, if the energy is initially given to a small packet of modes of very low frequency, the relaxation time to equilibrium (i.e. to equipartition) increases as a power of $1/\epsilon$ if $\epsilon > \epsilon^c$. Instead, if $\epsilon < \epsilon^c$, one has a first rapid (violent) relaxation to a natural packet extending up to a maximal frequency $\overline{\omega}(\epsilon) \simeq \epsilon^{1/4}$, while only on much larger time scales one would get equipartition. Moreover one starts now having analytical results confirming such a scenario, exactly in terms of the specific energy $\epsilon$ in the limit $N \to \infty$; such analytical results also give a precise analytical form for the spectrum corresponding to the natural packet mentioned above.

In conclusion, it seems that below a certain critical specific energy $\epsilon^c$ one deals with a metaequilibrium state, which is only later followed, on much longer time scales, by a real Maxwell–Boltzmann equilibrium. Such a scenario has many similarities with the one that is familiar for glasses and spin glasses, as was first pointed out in the paper [24].

## 4 The Landau–Teller Model of Molecular Collisions

We finally give a short review on the contribution of the internal degrees of freedom to the specific heat of polyatomic molecules.

It was recalled above that we owe to Boltzmann the fundamental idea that the contribution of the internal degrees of freedom to the specific heat would manifest itself only on long time scales, in contrast to the degrees of freedom of the center of mass (also called external degrees of freedom), which are known to relax to equilibrium after a few collisions. We also recalled how Jeans gave support to the point of view of Boltzmann, but later made a public retractation after the work of Poincaré on the necessity of quantization. It is also of interest to mention that the problem of the existence of long relaxation times was discussed from the experimental point of view at the first Solvay Conference, by Nernst and others. The opinion was there expressed that there was no evidence at all for such longer relaxation times.

Actually such longer relaxation times were observed experimentally in the year 1925, in studies of the dispersion (and anomalous diffusion) of sound in diatomic gases [25], and it turned out that the times were even 6 orders of magnitude larger than the mean collision time. In looking for an explanation of such longer relaxation times, in a period in which the discussions of Boltzmann and Jeans had been completely forgotten, a most relevant contribution was given by Landau and Teller [26]. They considered an extremely simplified model capturing the essence of the problem, namely the exchange of energy in a collision between an atom and a linear spring, with an exponential interatomic potential. For the energy exchange $\delta E$ they actually found an expression of a form already indicated by Jeans, namely $\delta E \simeq \exp\left(-\omega a/v\right)$, where $\omega$ is the frequency of the spring, $a$ the range of the potential and $v$ the velocity of the impinging atom. A relaxation time was then extracted from such a formula of the exchanged energy, and the common opinion was formed that the theory fits rather well the experimental data [27]. Serious doubts were however raised concerning the goodness of the agreement, as is witnessed for example by the following quotation [28]: "*It is impossible to determine whether the choice of the potential parameters is physically significant, because all errors in the theory are compensated by adjustable potential parameters*".

In any case, the common opinion is that the approach to equilibrium should be controlled by a single relaxation time, say $\tau_L$. Correspondingly, the law of temporal approach should be exponential [29], of the type $\exp(-t/\tau_L)$, as is familiar in the Onsager theory. We are of the opinion that the situation is here, however much more delicate. Indeed the Onsager theory is well suited for the approach to equilibrium of systems presenting a completely chaotic dynamics, while we are here confronted with the opposite situation, namely with systems that are nearly integrable, for which no one was able up to now to produce a statistical mechanics compatible with the dynamics (see however the papers [30], where indications are given that the statistics might be given according to the ideas of Einstein recalled above).

Further results were recently given in the paper [31]. There it is shown that the statistics induced by the dynamics in the Landau–Teller model of molecular collisions is a rather complex one, because over very long times the processes of the energy exchanges have many similarities to the well known Lévy processes, which are dominated by the presence of rare, but highly conspicuous, jumps. For an analogous situation in stellar dynamics see [32]. We are still working on the Landau–Teller model, and we hope to be able to show that here too one meets with two relaxation times, the shorter one leading to a state of metaequilibrium characterized by a "nonclassical" statistics, and the second one leading to the final Mazwell–Boltzmann equilibrium. This fact might be of interest for the phenomenology of sound dispersion.

## 5 Conclusions

We have recalled above how strong was the impact of the theorem of Poincaré concerning the necessity of quantization if the phenomenological law of Planck is to be recovered. In fact, after that work essentially all attempts at providing a classical understanding of Planck's law along the lines indicated by Boltzmann were abandoned. Peculiar exceptions were Einstein, who never proved convinced, and Nernst, who introduced the dynamical interpretation for the zero–point energy illustrated above.

Now, if one looks at the proof of Poincaré's theorem, it seems evident that the fundamental hypothesisis there made is that one should be dealing with a real equilibrium (i.e., almost by definition, with the Maxwell–Boltzmann distribution). And in fact on several occasions Poincaré had stressed that, if one recedes from equilibrium, one cannot have any thermodynamics at all: everything would become fuzzy. It seems difficult to disagree. Let us quote Poincaré himself [33]: "*Jeans tried to reconcile things, by supposing that what we observe is not a statistical equilibrium, but a kind of provisional equilibrium. It is difficult to take this point of view; his theory, being unable to foresee anything, is not contradicted by experience, but leaves without explanation all known laws*".

But in fact it seems to us that there is a possibility of getting a thermodynamics without a full equilibrium. The possibility is that one would be actually dealing, not just with a nonequilibrium, but rather with a situation of metaequilibrium, as for example in the case of glasses. In such a case one can have situations which apparently are indistinguishable from situations of a true equilibrium, so that the critique of Poincaré could be overcome. This is exactly what we are proposing, and is the reason why in this review we insisted in a particular strong way on the relevance of being able to findi at least two time scales in the problem of diatomic molecules (the existence ot two time scales in the Fermi–Pasta–Ulam problem being, by now, almost granted): only on an extremely long time scale would one get a true equilibrium, while on a first, short, time scale one would reach a metastable state (as in stellar dynamics). In turn, the termodynamics of the metastable state could not be described by the usual procedure of the type of Onsager, beacuse the latter makes reference to a chaotic dynamics, as stressed by Bowen, Ruelle, Sinai and by Gallavotti. In the state of metastability one is instead dealing, from a dynamical point of view, with the other extreme situation, namely with nearly integrable systems, and the suitable thermodynamics could perhaps be recovered along the lines suggested by Einstein and by Nernst. So, in conclusion, our suggestion is that, according to classical mechanics, Planck's law would describe a metaequilibrium state, at variance with quantum mechanics which interprets it as referring to a true equilibrium. It would be of a certain interest to ascertain whether the phenomenology might prove to be consistent with the scenario of metaequilibrium.

Finally we add a comment about Poincaré. It would appear that, if we are right in suggesting that the metaequilibrium scenario is a priori theoretically consistent, then Poincaré would be wrong. Thus we were very glad in discovering that Poincaré himself had some doubts about his attitude, essentially because he had to admit that, after all, there are more things in heaven and earth than his philosophy could imagine. Indeed, just a few lines after the destructive sentence concernig Jeans quoted above, in connection with the quantization of energy he added [33]: "*Will discontinuity reign over the physical universe and will its triumph be definitive? Or rather will it be recognised that such a discontinuity is only an appearence and that it dissimulates a series of continuous processes? The first person that saw a collision believed to be observing a discontinuous phenomenon, although we know today that the person was actually seeing the effect of very rapid changes of velocity, yet continuous ones*". It is true that Poincaré also adds the skeptical conclusion: "*To try to express today an opinion about these problems would mean to be wasting one's ink*". But at least we are comforted in learning that he admitted that other scenarios (such as that of Boltzmann, Einstein and Nernst, we would say) might be consistent.

## References


1. L. Galgani, A.Scotti: Phys. Rev. Lett. **28**, 1173 (1972)
2. C. Cercignani, L. Galgani, A.Scotti: Phys. Lett. A **38**, 403 (1972); L. Galgani and A. Scotti: *Recent progress in classical nonlinear dynamics*, Rivista Nuovo Cim. **2**, 189 (1972)
3. E. Fermi, J. Pasta and S. Ulam, in E. Fermi: Collected Papers (University of Chicago Press, Chicago, 1965); Lect. Appl. Math. **15**, 143 (1974)
4. A. Carati, L. Galgani: Found. Phys. **31**, 69 (2001)
5. D. Lynden Bell: Mon. Not. R. Astr. Soc. **136**, 101 (1967)
6. H. Poincaré, J. Phys. Th. Appl. **2**, 5 (1912), in *Oeuvres* IX, 626–653
7. *La théorie du rayonnement et les quanta*, ed. by P. Langevin and M. de Broglie (Gauthier–Villars, Paris 1912)
8. A. Einstein. In: *Albert Einstein: philosopher–scientist*, ed. by P.A. Schilpp (Tudor P.C., New York, 1949)
9. L. Boltzmann: *Lectures on Gas Theory*, transl. by S.G. Brush (University of California Press, Berkeley 1964); Nature **51**, 413 (1895)
10. J.H. Jeans: Phil. Mag. **35**, 279 (1903); Phil. Mag. **10**, 91 (1905)
11. Physics at the British Association: Nature **92**, 304 (1913); P.P. Ewald: *Bericht uber die Tagung der British Association in Birmingham (10-17 September)* Phys. Z. **14**, 1297 (1913), page 1298

12. N.O. Birge, S.R. Nagel: Phys. Rev. Lett. **54**, 3674 (1985); N.O. Birge: Phys. Rev. B **34**, 1631 (1986)
13. M. Doi, S.F. Edwards: *The Theory of Polymer Dynamics* (Clarendon Press, Oxford 1896)
14. W. Nernst: Verh. Dtsch. Phys. Ges, **18**, 83 (1916)
15. L. Galgani, G. Benettin: Lettere Nuovo Cim. **35**, 93 (1982); L. Galgani: Nuovo Cim. B **62**, 306 (1981); Lettere Nuovo Cim. **31**, 65 (1981); L.Galgani: in *Stochastic processes in classical and quantum systems*, ed. by S. Albeverio, G. Casati, D. Merlini, Lecture Notes in Physics N. 262 (Springer, Berlin 1986)
16. E. Fermi: Nuovo Cim. **25**, 267 (1923); Phys. Z. **24**, 261 (1923)
17. G. Benettin, G. Ferrari, L. Galgani, A. Giorgilli: Nuovo Cim. B **72** 137 (1982); G. Benettin, L. Galgani, A. Giorgilli: Poincaré's Non-Existence Theorem and Classical Perturbation Theory in Nearly-Integrable Hamiltonian Systems, in *Advances in Nonlinear Dynamics and Stochastic Processes*, ed. by R. Livi and A. Politi (World Scientific, Singapore 1985)
18. F.M. Izrailev and B.V. Chirikov: Sov. Phys. Dokl. **11**, 30 (1966)
19. P.Bocchieri, A.Scotti, B.Bearzi. A.Loinger: Phys. Rev. A **2**, 2013 (1970)
20. R. Livi, M. Pettini, S. Ruffo and A. Vulpiani: J. Stat. Phys. **48**, 539 (1987); D. Escande, H. Kantz, R. Livi, S. Ruffo: J. Stat. Phys. **76**, 605 (1994); D. Poggi, S. Ruffo, H. Kantz: Phys. Rev. E **52**, 307 (1995); J. De Luca, A.J. Lichtenberg, S. Ruffo: Phys. Rev. E **60**, 3781 (1999); L. Casetti, M. Cerruti–Sola, M. Modugno, G. Pettini, M. Pettini, R. Gatto: Rivista Nuovo Cim. **22**, 1 (1999); G. Parisi: Europhys. Lett. **40**, 357 (1997); A. Perronace, A. Tenenbaum: Phys. Rev. E **57** (1998) D.L. Shepelyansky: Nonlinearity **10**, 1331 (1997); P.R. Kramers, J.A. Biello, Y. Lvov: *Proceed. Fourth Int. Conf. on Dyn. Syst. and Diff. Eq., May 24–27, 2002, Wilmington, N.C.*, Discrete Cont. Dyn. Systems (in print)
21. L. Galgani, A. Giorgilli, A. Martinoli, S. Vanzini: Physica D **59**, 334 (1992)
22. G. Benettin, L. Galgani, A. Giorgilli: Nature **311**, 444 (1984); L. Galgani: in *Non-Linear Evolution and Chaotic Phenomena*, ed. by G. Gallavotti and P.F. Zweifel NATO ASI Series B: Vol. 176 (Plenum Press, New York 1988); G. Benettin, L. Galgani, A. Giorgilli: Comm. Math. Phys. **121**, 557 (1989); G. Benettin, L. Galgani, A. Giorgilli: Phys. Lett. A **120**, 23 (1987)
23. L. Berchialla, L. Galgani, A. Giorgilli: *Proceed. Fourth Int. Conf. on Dyn. Syst. and Diff. Eq., May 24–27, 2002, Wilmington, N.C.*, Discrete Cont. Dyn. Systems (in print)
24. A. Carati, L. Galgani: J. Stat, Phys. **94**, 859 (1999)
25. Pierce: Proc. Amer. Acad. **60**, 271 (1925); P.S.H. Henry, Nature **129**, 200 (1932)
26. L.D. Landau, E. Teller: Phy. Z. Sowjet. **10**, 34 (1936), in *Collected Papers of L.D. Landau*, ed. by ter Haar (Pergamon Press, Oxford 1965), page 147
27. K.F. Herzfeld, T.A. Litovitz: *Absorption and dispersion of ultrasonic waves* (Academic Press, New York and London, 1959); H.O. Kneser: in *Rendiconti della Scuola Internazionale di Fisica "Enrico Fermi": XXVII, Dispersion and absorption of sound by molecular processes* (Academic Press, New York and London, 1963); D. Rapp, T. Kassal: Chem. Rev. **64**, 61 (1969); A.B. Bhatia: *Ultrasonic Absorption* (Clarendon Press, Oxford, 1967); J.D. Lambert: *Vibrational and rotational relaxation in gases* (Clarendon Press, Oxford 1977); V.A. Krasilnikov, *Sound and ultrasound waves* (Moscow 1960, and Israel Program for Scientific Translations, Jerusalem 1963); H.O. Kneser: Schallabsorption und Dispersion in Gases, in *Handbuch der Physik XI–I* (Springer–Verlag, Berlin 1961)
28. D. Rapp, T. Kassal: Chem. Rev. **64**, 61 (1969)

29. O. Baldan, G. Benettin: J. Stat. Phys. **62**, 201 (1991); G. Benettin, A. Carati, P. Sempio: J. Stat. Phys. **73**, 175 (1993); G. Benettin, A. Carati, G. Gallavotti: Nonlinearity **10**, 479 (1997); G. Benettin, P. Hjorth, P. Sempio: J. Stat. Phys. **94**, 871 (1999)
30. A. Carati, L. Galgani: Phys. Rev. E **61**, 4791 (2000); A.Carati, L. Galgani: Physica A **280**, 105 (2000); A. Carati, L. Galgani: in *Chance in Physics*, ed. by J. Bricmont et al., Lecture Notes in Physics (Springer–Verlag, Berlin 2001)
31. A. Carati, L. Galgani, B. Pozzi: Phys. Rev. Lett. (2003, in print)
32. Dynamics and Thermodynamics of Systems with Long–Range Interactions, ed. by T.Dauxois, S Ruffo, E. Arimondo. M. Wilkens, Lecture Notes in Physics (Springer–Verlag, Berlin 2002)
33. H. Poincaré: Revue Scientifique **17**, 225 (1912), in *Oeuvres IX*, pp. 654–668

# Waves Derived from Galactic Orbits. Solitons and Breathers

Nikos Voglis

Academy of Athens, Research Center for Astronomy, 14 Anagnostopoulou Str., Athens, GR-10673, Greece

**Abstract.** We show how it is possible to define collections of non-interacting particles moving in the same potential so that solitons or breathers are formed. The displacements of particles in this case obey a partial differential equation (PDE). Thus it is possible to derive PDEs from a given Hamiltonian. We demonstrate the above methodology using, as an example, a cubic potential and we derive a Korteweg-de Vries equation.

We apply this methodology in galactic models and show that near resonances, where "Third Integrals" of motion can be defined, the motion of stars can be described in terms of solutions of a Sine-Gordon PDE. This equation admits kink or anti-kink solitons solutions.

In particular in the case of the Inner Lindblad Resonance (ILR), applying the Third Integral on a string of stars having as initial conditions the successive consequents of one orbit on a Poincaré surface of section a Frenkel-Kontorova Hamiltonian is constructed. The corresponding equations of motion are an infinite set of discrete Sine-Gordon equations. This set of equations admits solutions that represent localized oscillations on a grid that are known as Discrete Breathers. An analytic breather solution is derived and compared with the corresponding numerical solution in the case of a perturbed isochrone model.

The advantage of this methodology is that it takes into account the distribution of phases of stars moving under the same value of the third integral. Because of their nature, soliton solutions resist to dispersion and they can be a natural building block to construct more stable non-linear density waves in galaxies.

## 1 Introduction

Bars or spiral arms are of the main features in the internal structure of galaxies. Their stability is one of the most interesting problem in Galactic Dynamics. Efforts to explain this stability, particularly in the case of the spiral arms, give less stable structures than it is expected from the frequency of their observed occurrence. In reasonable galactic models spiral arms can survive for no more than several rotational periods (5 to 6).(See, for example, [19], [8], [17]).

As it is known, bars and spiral arms are maxima of density composed of different stars at different times, that is, they are maxima of density waves travelling along the azimuthal direction with pattern velocity which is different in general than the mean angular velocity of stars. This mechanism implies some phase correlations in the motion of stars in the galaxy. The main reason for which bars or spiral arms can be destroyed is because of the dispersion of velocities among stars that does not allow particular phase correlations to live for long.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 56–74, 2003.

A well known mechanism able to suppress the consequences of dispersion in dynamical systems composed of many particles is by the competition of the dispersion with the nonlinearity of the dynamical system. Such a competition is responsible, for example, for the robustness of solitons. If the spiral arms are formed by density waves of solitary nature one expects to have increased robustness compared with the linear waves. A special class of localized soliton-like solutions are the breathers (see, for example, [15]).

Solitons are solutions of nonlinear integrable Partial Differential Equations (PDEs). These solutions represent localized (not very extensive) waves travelling in a medium and preserving their identity (their basic parameters) even after interactions with other similar waves.

Nonlinearity appears in nature almost everywhere, while integrability is not very common. In most cases, nonlinearity leads to non-integrable dynamical systems. Non-integrability is characterized by chaos, i.e. a sensitive dependence on the initial conditions expressed by an exponential divergence of initially neighboring orbits. Thus, the question arises: Is non-integrability compatible with solitons?

Non-integrable systems contain both stable and unstable periodic orbits. Chaos appears only around the unstable periodic orbits. Around the stable periodic orbits the equations of motion are nearly integrable. It is worth looking for the possibility that solutions of integrable PDEs can describe the basic features of the behavior of dynamical system around stable periodic orbits.

The behavior of a dynamical system around stable periodic orbits is well known. Due to the KAM theorem, a stable periodic orbit is surrounded by invariant tori. Motion on these tori is regular, i.e. it resembles the motion in an integrable system. The respective integral, that keeps the motion on a torus, is the so called 'third integral' [5].

If the theory of solitons can describe the behavior of a dynamical system in the areas of regular motion, then this theory must be directly related to the theory of the third integral. Can we produce integrable PDEs and soliton solutions from the third integral?

In this paper we show how this can be obtained. We first give, in Sect. 2, an example to show how a sequence of non-interacting particles, moving in a potential $V(y)$, can be defined so that they form a soliton travelling along the direction $x$ on the $x-y$ plane.

We apply the same methodology in the epicyclic theory of the motion of stars in galactic potentials. Such potentials correspond to non-integrable Dynamical Systems. The motion of stars in galaxies can be either regular or chaotic, depending on the particular location in their phase space. The level of chaos, however, is low. We have found, for example, [20] that the Lyapunov Characteristic Numbers (LCN) of the orbits in galactic N-Body simulations is less than 0.1. This corresponds to a Lyapunov time (the time necessary for chaos to be effective) of the order more than 10 dynamical times. A black hole at the center of the galaxies can enhance the level of chaos, but even in the case of a huge black hole (e.g. of mass 1% of the total mass of the galaxy), only a very small number of stars describe orbits with LCN somehow larger than 0.1 [21].

In such an environment where order or weakly chaotic motion dominates it is possible to define nonlinear waves as we see below. In Sect. 3 we summarize the linear theory of the epicyclic motion and we show how a Klein-Gordon chain can be defined in terms of the Poincaré Surface Of Section (PSOS). In Sect. 4 we discuss in brief the notion of galactic resonances and the physical meaning of the 'slow angle'. Sections 3 and 4 serve mainly in giving the definitions and in understanding the role of various quantities in the problem. In Sect. 5 the nonlinear theory near the ILR in galactic models is discussed reviewing in brief the paper [6]. In this paper Contopoulos, using post-epicyclic approximation terms, derives a third integral $\Phi$ describing the motion of stars near the ILR. In Sect. 6 we show how a Sine-Gordon PDE and a corresponding Hamiltonian density can be written, based on $\Phi$. When $\Phi$ is applied for an infinite sequence of stars on the Poincaré S.O.S, a Frenkel-Kontorova Hamiltonian [10] can be constructed. The corresponding equations of motion are an infinite set of discrete Sine-Gordon equations which admit discrete soliton or discrete breather solutions. Analytic soliton or breather solutions are given in Sect. 7 and a comparison with a numerical application is given in Sect. 8. A summary and discussion is given in Sect. 9.

## 2 Solitons of Non-interacting Particles Moving in a Given Potential

In this section we show that it is possible to construct solitons of non-interacting particles that move in an 1-dimensional potential $V(y)$ on the x-y plane.

Consider, for example, the Hamiltonian

$$H = \frac{\dot{y}^2}{2c^2} - \frac{cy^2}{2} + \frac{y^3}{6} \tag{1}$$

where $c$ is a constant. For the value $H = 0$ of the Hamiltonian we get

$$\frac{dy}{y\sqrt{(1 - y/3c)}} = \pm c^{3/2} dt. \tag{2}$$

If we set

$$y = \frac{3c}{\cosh^2 a\phi} \tag{3}$$

we get

$$\phi = \phi_0 \pm \frac{c^{3/2}}{2a} t \tag{4}$$

where $\phi_0$ is the value of $\phi$ at $t = 0$.

On the x-y plane we consider a continuous sequence of non-interacting particles with coordinates $x = \phi_0$ and $y_0 = 3c/\cosh^2 ax$ and velocities $\dot{x} = 0$ and $\dot{y}_0 = \pm y_0\sqrt{1 - y_0/3c)}$. The sign (+) or (-) is chosen to be the sign of $x$. This sequence of particles forms a soliton travelling along the $x$ direction with velocity $v = c^{3/2}/2a = c$, for $2a = c^{1/2}$. This soliton is described by the equation

$$y = \frac{3c}{\cosh^2 \frac{\sqrt{c}}{2}(x - ct)} \tag{5}$$

and it can be derived as a particular solution

$$u = y(x - ct) = y(\phi) \tag{6}$$

of the well known Korteweg-de Vries PDE

$$\frac{\partial u}{\partial t} + u\frac{\partial u}{\partial x} + \frac{\partial^3 u}{\partial x^3} = 0 \tag{7}$$

Notice that, under the transformation (6) the above PDE can be integrated twice to give

$$\frac{y_\phi^2}{2} - c\frac{y^2}{2} + \frac{y^3}{6} = Ay + B \tag{8}$$

where A and B are the constants of integrations. In the particular case of $A = 0$ at a fixed value of $x$ we can replace $y_\phi$ by $y_\phi = -\dot{y}/c$ to find the Hamiltonian (1).

## 3 Linear Epicyclic Waves in Galactic Models

The idea of forming waves with collections of non-interacting particles moving in the same potential can be applied in galactic models. In this section we show how this can be obtained in the simple case of the linear epicyclic motion.

In a rotating galactic model (either bared or spiral) with a plane of symmetry $(r, \theta)$ in polar coordinates the motion of stars on this plane can be given by the Hamiltonian

$$H = \frac{1}{2}(\dot{r}^2 + \frac{J_0^2}{r^2}) - \Omega_s J_0 + V_0(r) + V_1(r, \theta) = H_0 + V_1(r, \theta) \tag{9}$$

where $J_0$ is the angular momentum of a star with respect to an inertial frame, $V_0(r)$ is the axisymmetric component of the potential, $V_1(r, \theta)$ is the perturbation of the potential due either to the bar of the galaxy or to the spiral arms that form a pattern rotating with angular velocity $\Omega_s$. $H_0$ is the unperturbed (axisymmetric) part of the Hamiltonian.

We define a cartesian coordinate system XY rotating with the pattern, i.e. with angular velocity $\Omega_s$. The polar angle $\theta$ is measured in the frame XY from the X-axis.

In the unperturbed potential $V_0(r)$ a star of $H_0 = h$ (the Jacobi integral) and angular momentum $J_0 = J_c$ describes a circular orbit with radius $r_c$ such that

$$h = \frac{J_c^2}{2r_c^2} - \Omega_s J_c + V_0(r_c) \tag{10}$$

where $r_c$ is the root of the equation

$$J_c^2 = r_c^3 \frac{dV_0(r_c)}{dr_c} \tag{11}$$

We define the radial action $J_1$ as

$$J_1 = \frac{1}{2\pi} \oint \dot{r} dr \tag{12}$$

and the azimuthal action $J_2$ as $J_2 = J_c$. If the unperturbed part $H_0$ of the Hamiltonian (9) is expanded up to the linear term in the radial action in Taylor series around the radius $r_c$, it becomes

$$H_0 = \frac{J_2^2}{2r_c^2} - \Omega_s J_2 + V_0(r_c) + \omega_1 J_1 \tag{13}$$

where $\omega_1$ is the epicyclic frequency given by

$$\omega_1^2 = V_0''(r_c) + \frac{3J_2^2}{r_c^4} \tag{14}$$

(a prime denotes the derivative with respect to $r$).

In terms of the radial epicyclic coordinate $x = r - r_c$ and its conjugate momentum $p$ the radial action is

$$J_1 = \frac{1}{2}(\omega_1 x^2 + \frac{p^2}{\omega_1}) \tag{15}$$

The guiding center (i.e. the center of the epicycle) describes the circular orbit $r_c$ with respect to an inertial frame with angular velocity

$$\Omega_c = \frac{J_2}{r_c^2} \tag{16}$$

while with respect to the rotating XY frame with angular velocity

$$\omega_2 = \Omega_c - \Omega_s \tag{17}$$

The epicyclic angle $\theta_1$ and the azimuthal angle $\theta_2$ conjugate to $J_1$ and $J_2$ respectively, are determined by the equations

$$\dot{\theta}_1 = \frac{\partial H_0}{\partial J_1} = \omega_1 \ , \qquad \dot{\theta}_2 = \frac{\partial H_0}{\partial J_2} = \omega_2 \tag{18}$$

from which we get

$$\theta_1 = \omega_1(t - t_a) \ , \qquad \theta_2 = \omega_2 t \tag{19}$$

where $t_a$ is a constant.

The azimuthal angle $\theta_2$ gives the position of the guiding center of the epicycle. We can define the origin of $\theta_2$ so that the guiding center crosses the X-axis at $t = 0$.

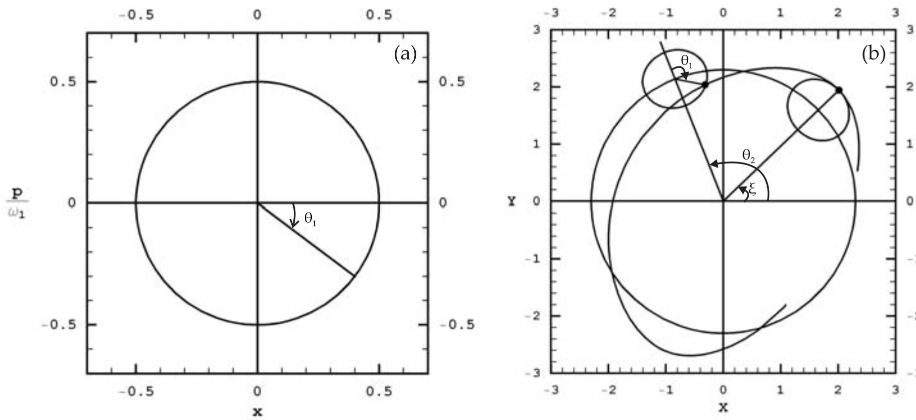



**Fig. 1.** (a) The phase space $(x, p/\omega_1)$ of the epicyclic motion and the definition of the epicyclic angle $\theta_1$. (b) The azimuthal angle $\theta_2$ gives the position of the guiding center. The angle $\xi$ is the value of $\theta_2$ when the star is a its apocenter.

The angle $\theta_1$ is the phase angle on the phase space $(x, p/\omega_1)$, (Fig.1a). This angle measures the phase of the star on its orbit. For example, if we define $\theta_1$ so that $\theta_1 = 0$ when the star is at its apocenter, then, when $\theta_1 = \pi$, the star is at its pericenter. All other values of $\theta_1$ cover all the intermediate phases. Under these definitions the constant $t_a$ in (19) is the time when the star reaches its apocenter.

If $\xi$ is the value of $\theta_2$ at the time when the star is at a particular phase of its orbit, e.g. at its apocenter, then $\xi = \omega_2 t_a$(Fig.1b). The epicyclic angle $\theta_1$ is written as

$$\theta_1 = \omega_1 t - k\xi \tag{20}$$

where $k = \frac{\omega_1}{\omega_2}$. The epicyclic coordinate $x$ and its conjugate momentum $p$ are expressed in terms of the action $J_1$ and the angle $\theta_1$ as

$$x = \sqrt{\frac{2J_1}{\omega_1}} \cos\theta_1 = \sqrt{\frac{2J_1}{\omega_1}} \cos\left(\omega_1 t - k\xi\right) , \tag{21a}$$

$$\frac{p}{\omega_1} = -\sqrt{\frac{2J_1}{\omega_1}} \sin\theta_1 = -\sqrt{\frac{2J_1}{\omega_1}} \sin\left(\omega_1 t - k\xi\right) . \tag{21b}$$

The angle $\xi$ can be used as an independent variable. Let us consider a continuous sequence of stars with the same actions $J_1$ and $J_2$, uniformly distributed along the values of $\xi$ at a given time. All these stars follow the same orbit but with different initial phases. According to (21a), their motions (in and out the circular orbit $r_c$) form a continuous sinusoidal wave, with wavenumber $k$, travelling along the $\xi$-axis.

We can easily show that (21a) is a solution of the Klein-Gordon wave equation

$$x_{tt} + \omega_2^2 x_{\xi\xi} + 2\omega_1^2 x = 0 \tag{22}$$

The term $\omega_2^2 x_{\xi\xi}$ in this equation expresses a kind of phase interaction between neighboring stars in the sequence.

Consider the Poincaré surface of section (PSOS) of the orbit of a single star, with actions $J_1$ and $J_2$, when for example $\theta_2 = 0$, that is, when the star crosses the positive X-axis with a positive velocity component along the Y-axis ($\dot{y} > 0$). If the star starts from its apocenter that is initially on the above PSOS then at $t = 0$, $\theta_1 = 0$ and $\theta_2 = 0$. This means that the value of $\xi$ for this star is $\xi = 0$.

For irrational values of $k = \omega_1/\omega_2$, the successive consequents of this orbit on the PSOS, that occur in times $t_n = 2\pi n/\omega_2$ with $n = 0, \pm 1, \pm 2, ...$, belong to an invariant curve of action $J_1$ on the plane $(x, p/\omega_1)$. The rotation angle of these consequents on the invariant curve is $\theta_1$ and takes the values

$$\theta_{1n} = \omega_1 t_n = 2\pi n k \tag{23}$$

Consider now an infinite sequence of stars numbered as $n = 0, \pm 1, \pm 2, ...$ with the same actions $J_1$, $J_2$, having as initial conditions (at $t = 0$) the above consequents. From (20) and (23) we get

$$\xi_n = -2\pi n \tag{24}$$

Therefore the epicyclic angles of these stars as functions of time are

$$\theta_{1n}(t) = \omega_1 t + 2\pi n k \tag{25}$$

and their equations (21a and 21b) become

$$x_n = \sqrt{\tfrac{2J_1}{\omega_1}} \cos\left(\omega_1 t + k 2\pi n\right) , \tag{26a}$$

$$\frac{p_n}{\omega_1} = -\sqrt{\tfrac{2J_1}{\omega_1}} \sin\left(\omega_1 t + k 2\pi n\right) . \tag{26b}$$

Equation (26a) describes a discrete sinusoidal wave travelling along an 1-dimensional grid of particles. This equation is a solution of the discrete set of Klein-Gordon ordinary differential equations

$$\ddot{x}_n + \frac{\omega_2^2}{4\pi^2}(x_{n+1} + x_{n-1} - 2x_n) + \omega^2 x_n = 0 \tag{27}$$

where $\omega$ is given by the equation

$$\omega^2 = \omega_1^2 \left(1 + \frac{\sin^2 \frac{\omega}{\omega_2}\pi}{(\frac{\omega}{\omega_2}\pi)^2}\right) \tag{28}$$

that tends to the value of $\omega = \sqrt{2}\omega_1$ provided that the wavenumber $k = \frac{\omega}{\omega_2}$ satisfies the condition $k \ll 1$.

It is remarkable that the second term in (27) represents a phase interaction between neighboring stars in the sequence, as if this interaction were due to repulsing elastic forces.

Thus, up to now, we have shown that using the invariance of the epicyclic action $J_1$ and the azimuthal action $J_2$, we can define continuous epicyclic waves satisfying a Klein-Gordon equation (22). Furthermore, in terms of the Poincaré S.O.S., we can construct a Klein-Gordon chain of stars satisfying a corresponding set of ordinary differential equations (27).

The above theory can be generalized in the non-linear case as it will be described in Sect. 5.

# 4 Resonances in the Unperturbed Potential. The Inner Lindblad Resonance and the Slow Angle $\psi$

In a central potential $V_0(r)$ the apocenters of an orbit in general precess. The radial frequency $\Omega_r$ and the mean azimuthal frequency $\Omega_a$ of an orbit are defined as

$$\Omega_r = \frac{2\pi}{T_r}, \qquad \Omega_a = \frac{\phi}{T_r} \tag{29}$$

where $T_r$ is the time from one apocenter to the next, called radial period, and $\phi$ is the azimuthal angle between two successive apocenters.

In general this orbits is not closed, i.e. is not periodic. However, this orbit closes, if it is observed in a frame rotating with angular velocity

$$\Omega_s = \frac{\phi - 2\pi}{T_r} = \Omega_a - \Omega_r \tag{30}$$

In other words it becomes periodic resembling in this case a Keplerian ellipse. Such a closed orbit is called 1:1 resonant periodic orbit.

If the frame rotates with angular velocity

$$\Omega_s = \frac{2\phi - 2\pi}{2T_r} = \Omega_a - \frac{\Omega_r}{2} \tag{31}$$

the orbit closes again but in this case it resembles an ellipse symmetric with respect to the center. This is a 2:1 resonant periodic orbit. Equation (31) in terms of the frequencies $\omega_1 = \Omega_r$ and $\omega_2 = \Omega_a - \Omega_s$ used in the previous section can be written as

$$\omega_1 = 2\omega_2 \tag{32}$$

(Notice that the mean azimuthal frequency $\Omega_a$ is identical to the angular velocity $\Omega_c$ of the guiding center). Equation (32) expresses the well known Inner Lindblad Resonance (ILR) in galaxies.

Consider a 2:1 resonant orbit, i.e. an orbit exactly at the ILR of a galaxy in the XY plane in the unperturbed potential $V_0(r)$. The angle between the major axis of the orbit and the X-axis is exactly the angle $\xi$, defined in the previous section that determines the orientation of the apocenter. This orbit can be analyzed in circular and epicyclic motion as in the previous section.

The slow angle $\psi$ (introduced by Lynden-Bell [13]) is defined as

$$\psi = \theta_1 - 2\theta_2 \tag{33}$$

Using (19) and (20) we get

$$\psi = (\omega_1 - 2\omega_2)t - 2\xi \tag{34}$$

Since the orbit is exactly at the resonance ($\omega_1 - 2\omega_2 = 0$), we have

$$\psi = -2\xi \tag{35}$$

This means that the angle $\psi$ for this orbit is constant and determines the orientation of the major axis of the orbit. For other orbits, however, near to the resonance, the angle $\psi$ varies slowly with $\dot{\psi}$ measuring the slow precession of the major axis of the orbit. For this reason the angle $\psi$ is called slow angle or precession angle [13], [6], [14], [16].

On the PSOS as defined in the previous section, i.e. when $\theta_2 = 0$ the values of the angle $\psi$ coincide with the corresponding angle of $\theta_{1n}$ of (23). Thus the angle $\psi$ on the PSOS $(x, p/\omega_1)$ plays the role of the rotation angle on an invariant curve.

## 5 Non-linear Epicyclic Motion near the Inner Lindblad Resonance

Any perturbation $V_1(r,\theta)$ imposed on the axisymmetric potential $V_0(r)$ can be analyzed in Fourier modes [11] as

$$V_1(r,\theta_1,\theta_2) = \epsilon \sum_{lm} V_{lm}(J_1,J_2)\cos(l\theta_1 - m\theta_2) \tag{36}$$

where $\epsilon$ is a small positive quantity and $l, m$ are integers. For the orbits near the ILR the most important term of the Fourier modes in the expansion (36) (i.e. the mode that can pump more energy on such orbits than all other terms) is the term with $l = 1$ and $m = 2$, namely, the term

$$\epsilon V_{12}(J_1,J_2)\cos(\theta_1 - 2\theta_2) = \epsilon V_{12}(J_1,J_2)\cos\psi \tag{37}$$

This is the resonant term in the ILR.

Using post epicyclic approximations Contopoulos [6] has shown that the Hamiltonian (9) can be written in the form

$$H = h + \omega_1 I_1 + \omega_2 I_2 + aI_1^2 + 2bI_1I_2 + cI_2^2 + ... + V_1(r,\theta) \tag{38}$$

where

$$I_1 = \frac{1}{2\pi}\oint \dot{r}dr = J_1 \tag{39}$$

is the radial action exactly the action $J_1$ defined in Sect. 2.

The azimuthal action $I_2$ in the Hamiltonian (38) is the excess from the angular momentum of the circular orbit

$$I_2 = J_0 - J_c \tag{40}$$

which is of the same order as $J_1$. The frequencies $\omega_1$ and $\omega_2$ have the same meaning as in Sect. 2.

Contopoulos [6] shows that near the ILR under proper canonical transformations the Hamiltonian (38) leads to an equivalent Hamiltonian $\mathcal{H}$ up to terms of $O(\epsilon)$

$$\mathcal{H} = \omega_2 J_2 + cJ_2^2 + \Phi(J_1,\psi) = 0 \tag{41}$$

where $\Phi(J_1, \psi)$ is a third integral given by

$$\Phi(J_1, \psi) = \gamma J_1 + \alpha J_1^2 + \epsilon_1 (\frac{2J_1}{\omega_1})^{1/2} (J_{20} - 2J_1) \cos \psi \tag{42}$$

describing the epicyclic motion near the inner Lindblad resonance in post epicyclic approximation terms. In these expressions $\epsilon_1$ is a small positive constant proportional to $\epsilon$ in (37), $\gamma$ is defined as

$$\gamma = \omega_1 - 2\omega_2 \tag{43}$$

and $\alpha$, $c$, $J_{20}$ are constants.

The azimuthal action is $J_2 = I_2 + 2J_1$, the so called fast action. Its conjugate fast angle $\psi_2$ (equal to the azimuthal angle $\theta_2$)is ignorable in the Hamiltonian (41). Therefore the azimuthal action $J_2$ is a constant and the problem up to terms of $O(\epsilon)$ is integrable.

The angle $\psi$ is the slow angle given by (33). Its conjugate action is $J_1$ that can be written as

$$J_1 = \frac{1}{2} (\frac{\overline{p}^2}{\omega_1} + \omega_1 \overline{x}^2) \ . \tag{44}$$

in terms of the epicyclic coordinate

$$\overline{x} = (\frac{2J_1}{\omega_1})^{1/2} \cos \psi \tag{45}$$

and its conjugate momentum

$$\frac{\overline{p}}{\omega_1} = -(\frac{2J_1}{\omega_1})^{1/2} \sin \psi \ . \tag{46}$$

In terms of $\overline{x}$ and $\overline{p}$ the third integral (42) is a fourth order polynomial. A periodic orbit of the system is located at an extreme value of $\Phi(J_1, \psi)$. In other words periodic orbits are found as the roots of algebraic system

$$\Phi_x = 0, \qquad \Phi_p = 0 \tag{47}$$

where the index denotes the corresponding partial derivative. A root of this system gives a stable or an unstable periodic orbit depending on the sign of the quantity $S = \Phi_{xx}\Phi_{pp} - \Phi_{xp}^2$ (stable for $S > 0$).

For very small values of the perturbation parameter $\epsilon_1$ the system (47) has only one real root giving a stable periodic orbit called $x_1$, while for larger values of $\epsilon_1$ three roots appear, by a saddle point bifurcation, giving two stable orbits called $x_1$ and $x_2$ and one unstable orbit called $x_3$ (Figs.1a,b in [6]).

## 6 A Sine-Gordon Equation

The equations of motion derived from $\mathcal{H}$ in (41) give

$$\dot{\psi}_2 = \dot{\theta}_2 = w = \frac{\partial \mathcal{H}}{\partial J_2} = \omega_2 + 2cJ_2 = constant \tag{48}$$

$$\dot{\psi} = \frac{\partial \mathcal{H}}{\partial J_1} = \frac{\partial \Phi}{\partial J_1} = \gamma + 2\alpha J_1 + \epsilon_1 A' \cos\psi \tag{49}$$

$$\dot{J}_1 = -\frac{\partial \mathcal{H}}{\partial \psi} = -\frac{\partial \Phi}{\partial \psi} = \epsilon_1 A(J_1) \sin\psi \tag{50}$$

where

$$A = (\frac{2J_1}{\omega_1})^{1/2}(J_{20} - 2J_1) \tag{51}$$

and $A'$ is the derivative of $A$ with respect to $J_1$.

The second time derivative of $\psi$ obeys the pendulum equation up to $O(\epsilon_1)$ terms, i.e.

$$\ddot{\psi} - \omega_0^2 \sin\psi = 0 \tag{52}$$

where

$$\omega_0^2 = \epsilon_1[2\alpha A(J_1) - (\gamma + 2\alpha J_1)A'] \tag{53}$$

A pendulum equation similar to (52) is known in Galactic Dynamics [16] derived by alternative arguments. It expresses the possibility that an orbit precesses so that its major axis either librates around a fixed direction at $\psi = \pi$, or rotates around the center of the galaxy.

In the case when the three periodic orbits $x_1$,$x_2$,$x_3$ appear, one can find both librating or rotating orbits with a separatrix between them passing through the unstable orbits $x_3$.

If (52) is integrated once we get

$$\frac{\dot{\psi}^2}{2} = -\omega_0^2 cos\psi + C \tag{54}$$

where $C$ is the integration constant. This constant is a measure of the value of the third integral $\Phi(J_1, \psi)$ at a given orbit. It can be easily shown that, up to terms of $O(\epsilon)$,

$$\gamma^2 + 4\alpha\Phi = 2C \tag{55}$$

In the general case (for any value of C) (54) can be further integrated in term of the Jacobian elliptic functions (e.g. see [1]). On the separatrix, $\psi$ reaches the value $\psi = 0$ with $\dot{\psi} = 0$, thus we get $C = \omega_0^2$. For this value of C the solution of (54) is

$$\tan\frac{\psi}{4} = e^{\pm[\omega_0(t - t_0)]} \tag{56}$$

where $t_0$ is the time when $\psi = \pi$. On the PSOS, according to (45), this value of $\psi$ corresponds to the pericenter of the orbit, since $\overline{x}$ becomes minimum. The value $\psi = 0$, on the other hand, corresponds to the apocenter of the orbit. Such an orientation, of an orbit, on the separatrix, is approached as time goes to infinity.

Introducing the angle $\xi$ as in the previous sections

$$\xi = wt_0, \qquad k_0 = \frac{\omega_0}{w} \tag{57}$$

$$\tan\frac{\psi}{4} = e^{\pm(\omega_0 t - k_0\xi)} \tag{58}$$

We consider an continuous sequence of stars with the same constant $C = \omega_0^2$ (on the separatrix) with characteristic phase angles $\phi = \omega_0 t - k_0\xi$. Equation (58) represents a kink or antikink soliton travelling along $\xi$ and is a solution of the Sine-Gordon wave equation

$$\ddot{\psi} + w^2\psi_{\xi\xi} - 2\omega_0^2 \sin\psi = 0 \tag{59}$$

The Lagrangian density function from which this equation can be derived is

$$\mathcal{L} = \frac{\dot{\psi}^2}{2} + \frac{w^2\psi_\xi^2}{2} - 2\omega_0^2\cos\psi \tag{60}$$

and the corresponding Hamiltonian density

$$\Phi_* = \dot{\psi}\frac{\partial L}{\partial\dot{\psi}} - \mathcal{L} = \frac{\dot{\psi}^2}{2} - \frac{w^2\psi_\xi^2}{2} + 2\omega_0^2\cos\psi \tag{61}$$

It is easy to show that this Hamiltonian density can be expressed up to $O(\epsilon)$ in terms of the constant $C$ in (54) or the third integral $\Phi$ in (42) as

$$\Phi_* = 2C - \frac{1}{2}(\dot{\psi}^2 + w^2\psi_\xi^2) = \gamma^2 + 4\alpha\Phi - \frac{1}{2}(\dot{\psi}^2 + w^2\psi_\xi^2) \tag{62}$$

If we consider an infinite number of stars starting at $t = 0$ with $\xi = -2\pi n$, $n = 0, \pm1, \pm2, ...$, i.e. on the separatrix of the Poincaré S.O.S. as described in the previous sections, (58) becomes

$$\tan\frac{\psi_n}{4} = e^{\pm(\omega_0 t + k_0 2\pi n)} \tag{63}$$

Since $k_0 \ll 1$ this is a solution of the discrete set of the ordinary Sine-Gordon equations

$$\ddot{\psi}_n + \frac{w^2}{4\pi^2}(\psi_{n+1} + \psi_{n-1} - 2\psi_n) - 2\omega_0^2\sin\psi_n = 0 \tag{64}$$

These equations can also be derived directly from a Fenkel-Kontorova Hamiltonian constructed in terms of the Hamiltonian density (61) by replacing the partial derivative $\psi_\xi$ by its discrete equivalent and sum over all the $n$ stars with initial conditions on a single invariant curve of the Poincaré S.O.S. (same value of $\Phi_*$). Thus we get

$$\Phi_{FK} = \sum_n \Phi_{*n} = \sum_n [\frac{\dot{\psi}_n^2}{2} - \frac{w^2}{8\pi^2}(\psi_{n+1} - \psi_n)^2 + 2\omega_0^2\cos\psi_n] \tag{65}$$

If $\psi_n$ is a periodic function with frequency $\omega$ and amplitude $X_n$, i.e. $\psi_n = X_n \exp\omega t$, the dynamics of $\psi_n$ in (64) can be studied through the following map produced by direct substitution in (64)

$$\begin{aligned} Y_{n+1} &= Y_n + K\sin X_n + \Lambda X_n\ , \\ X_{n+1} &= X_n + Y_{n+1}\ , \end{aligned} \tag{66}$$

where $K = 2(\omega_0 T)^2$ and $\Lambda = (\omega T)^2$, with $T = 2\pi/w$.

## 7 Analytic Solutions

If we make the transformation

$$\phi = \omega_0 t - k_0 \xi \tag{67}$$

the Sine-Gordon equation (59) after a first integration becomes

$$\psi_\phi^2 = 2[c + 1 - 2\sin^2 \frac{\psi + \pi}{2}] \tag{68}$$

where $c$ is the integration constant and $c \geq -1$. The solution of this equation is given by the Jacobian sn-oidal Elliptic Functions. Namely, if

$$m = (c+1)/2 \tag{69}$$

the solution is given by

$$\cos\frac{\psi}{2} = m^{1/2} sn(\phi \mid m) \tag{70}$$

This solution for $\psi$ is a soliton travelling along $\xi$ with velocity $d\xi/dt = k_0/\omega_0 = w$. Using (33) and (48) the epicyclic coordinate $x = x_m \cos\theta_1$ can be expressed as

$$x = x_m \cos(\psi + 2wt) = x_m(\cos\psi\cos 2wt - \sin\psi \sin 2wt) \tag{71}$$

where $x_m = \sqrt{2J_1/\omega_1}$ is the amplitude of the epicyclic coordinate $x$. This amplitude varies since the action $J_1$ is a function of time according to (50). Integrating (50) we finally get

$$x_m = x_0 \pm \beta m^{1/2} \sqrt{1 - sn^2(\phi \mid m)} \tag{72}$$

where $x_0$ is the epicyclic coordinate for $\psi = 0$ and $\beta$ measures the amplitude of variations of $x_m$ around the value of $x_0$. The proper choice of sign in (72) depends on the sign of $\dot\psi$ (minus for $\dot\psi < 0$).

From (70) we have

$$\cos\psi = 2m sn^2(\phi \mid m) - 1 \tag{73}$$

$$\sin\psi = 2m^{1/2} sn(\phi \mid m)\sqrt{1 - m sn^2(\phi \mid m)} \tag{74}$$

In the case of $m \ll 1$ we get the harmonic oscillation limit

$$\cos\psi = 2m^{1/2}\sin^2\phi - 1 \tag{75}$$

$$\sin\psi = 2m^{1/2}\sin\phi(1 - \frac{m}{2}\sin^2\phi) \tag{76}$$

In the particular case of $m = 1$ the motion occurs on one of the two separatrices. In this case the expressions (73) and (74) become

$$\cos\psi = 1 - \frac{2}{\cosh^2\phi}, \qquad \sin\psi = \frac{2\tanh\phi}{\cosh\phi}\ . \tag{77}$$

Let $R_g$ be the radius of the circle described by the guiding center. The radius $R_3$ to a star describing the unstable periodic orbit $x_3$ can be approximated by

$$R_3 = R_g + x_0 \cos 2wt \tag{78}$$

The radius to any star describing an epicyclic motion with the same guiding center can be written as

$$R = R_g + x_m \cos \theta_1 \tag{79}$$

For the stars moving on the separatrix surrounding the periodic orbits $x_2$ (where $\dot{\psi} < 0$) the correct sign in (72) is (-). In this case the difference $\Delta R = R - R_3 = x_m \cos \theta_1 - x_0 \cos 2wt$ can be written as

$$\Delta R = [-2(x_0 - \frac{\beta}{\cosh \phi})\frac{1}{\cosh^2 \phi} - \frac{\beta}{\cosh \phi}] \cos 2wt$$

$$+2(x_0 - \frac{\beta}{\cosh \phi})\frac{\tanh \phi}{\cosh \phi} \sin 2wt \tag{80}$$

The r.h.s in (80) represents a breather (composed of two simpler superposed breathers). This is a breather of a continuous sequence of stars along $\xi$. For the discrete sequence $\{\mathbf{n}\}$ of stars the angle $\phi$ is

$$\phi = \omega_0 t + k_0 2\pi n \tag{81}$$

In this case we have a discrete breather on an one-dimensional grid of stars, similar to breathers found in other branches of Physics. Studies on discrete breathers is currently a very active field of research (see e.g. [18], [22], [2], [15], [9], [3], [4], [33], [23], see also Bountis and Bergamin in this volume). One-dimensional discrete breathers are defined as common frequency oscillations of a not very large number of particles localized in a region of an infinite grid of interacting particles. The localization is due to the fact that the amplitudes of the oscillations decrease exponentially with the distance from a given point of the grid due to the nonlinearity of the potential. Discrete breathers can be either stationary, remaining at the same region of the grid or they can travel along the grid. In the present case the breather travels along $-n$ with a phase velocity $d\xi/dt = -2\pi \Delta n/\Delta t = \omega_0/k_0 = w$.

## 8 A Numerical Application and Comparison with the Analytic Results

We present below a comparison of the above solution (80) with the numerical calculations in a particular galactic model. In this model the axisymmetric component of the potential is taken to be the well known isochrone model

$$V_0(r) = -\frac{1}{1 + \sqrt{1 + r^2}} , \tag{82}$$

on which a bar-like perturbation

$$V_1(r, \theta) = \epsilon r^{1/2}(16 - r) \cos 2\theta \qquad (83)$$

is superposed. This model has been used in [7] in studying galactic orbits in week and strong bars. The adopted values of $\epsilon$ and $\Omega_s$ are $\epsilon = 0.00001$ and $\Omega_s = 0.05$. For a value of the Jacobi constant $h = -0.28$, motion occurs near the ILR. The three periodic orbits $x_1$, $x_2$ (stable) and $x_3$ (unstable) on the XY plane are shown in Fig. 2, while the corresponding Poincaré S.O.S. $(X, \dot{X})$ is shown in Fig.3. In this figure the axes are $X = R_g + x$ and $\dot{X} = \dot{x}$.

This system is non-integrable. For the above choice of parameters, chaos in the region of the unstable orbit $x_3$ at $(X_3 \approx 2.33,\ \dot{X} = 0)$ is very small, so that it can be neglected. The stable periodic orbits $x_1$ and $x_2$ are at the centers of the two islands at $(X \approx 1.03,\ \dot{X} = 0)$ and $(X \approx 1.82,\ \dot{X} = 0)$, respectively. These islands are limited by two separatrices. (In fact, what we call a separatrix here, is a very thin homoclinic tangle formed by the unstable and the stable asymptotic curves emanating from the unstable periodic orbit $x_3$). The dots in Fig.3 correspond to the successive consequents of a single orbit on each separatrix.

Let us focus on the points of one separatrix only, for example, the separatrix surrounding the periodic orbit $x_2$. We number these consequents as $n = 0, \pm 1, \pm 2, ...$ composing the sequence $\{\mathbf{n}\}$ of stars. The consequent $n = 0$ is at $\dot{X} \approx 0$ and $X_{min} = 1.27$. Positive or negative $n$ means a consequent in the future or in the past, respectively, relative to the consequent at $n = 0$.

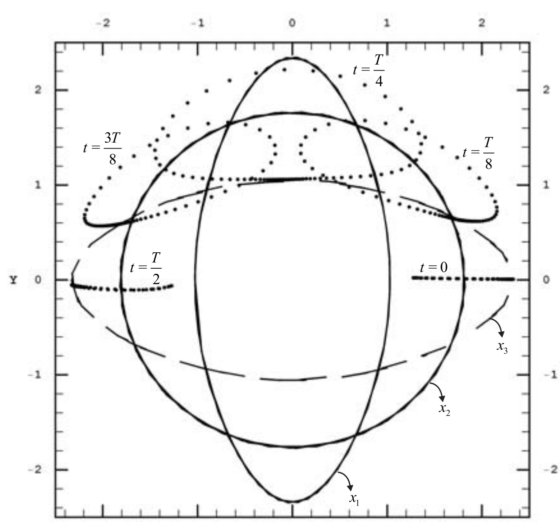



**Fig. 2.** The three periodic orbits $x_1$,$x_2$ (stable) and $x_3$ (unstable) on the XY plane. The chain of stars forming the breather of Fig.4 is shown at several snapshots.

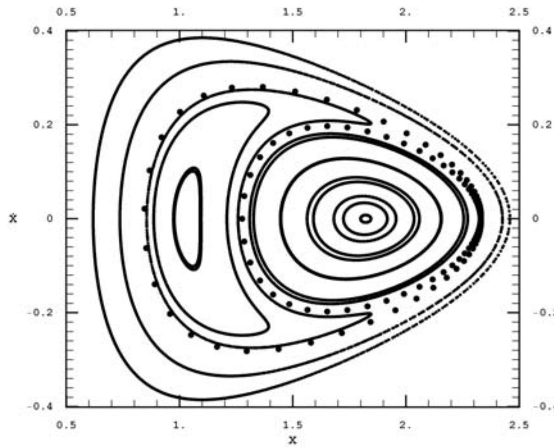



**Fig. 3.** The Poincaré surface of section in the model of (82),(83) for $h = -0.28$. The three periodic orbits $x_1$,$x_2$, and $x_3$ are at $X = 1.03$, $X = 1.82$, $X = 2.33$, respectively, with $\dot{X} = 0$. Dots represent the successive consequents of an orbit starting on each separatrix. The dots on the separatrix surrounding $x_2$ are the initial conditions of the chain of stars forming the breather shown in Fig.4.

We run simultaneously the orbits of a chain of $2n_{max} + 1 = 145$ stars with initial conditions the above consequents and we calculated

$$\Delta R(n,t) = R_n(t) - R_{x_3}(t) \tag{84}$$

where $R_n(t)$ is the radial distance from the center at time $t$ of the star $n$ and $R_{x_3}(t)$ is the same as $R_n(t)$, but of a star moving exactly on the unstable periodic orbit $x_3$.

In Fig. 4 the line with dots gives $\Delta R(n,t)$ as a function of $n$ at four snapshots at times $t = 0,\ T/8,\ 2T/8,\ 3T/8$, where $T$ is the azimuthal period corresponding to the frequency $w$. The thin solid line gives at the same times $\Delta R$ as a function of $\xi$ evaluated from (80). The values of $x_0$ and $\beta$ are taken from the data of Fig. 3. They are $x_0 = 0.65$ and $\beta = 0.24$. The agreement between the numerical $\Delta R(n,t)$ and the analytic $\Delta R$ is quite good. The small differences are mainly due to the approximate representation of the orbit $x_3$ by the equation (78) but also to the fact that the analytic solution contains only $\mathrm{O}(\epsilon)$ terms. Similar results are found if we consider the other separatrix of Fig.3.

We see therefore that proper chains of stars can form breathers travelling along the chain with a speed w equal to the angular velocity of the guiding center. The chain of stars used in the numerical breather of Fig.4 is shown in real space in Fig.2 at $t = 0,\ T/8,\ 2T/8,\ 3T/8, T/2$. The stars of this chain lie initially ($t = 0$) on a straight line along the $X$ axis. The chain evolves to a loop at $T/4$ and forms again an almost straight line along $X$ at $T/2$ and so on, being successively inside or outside the unstable periodic orbit $x_3$.

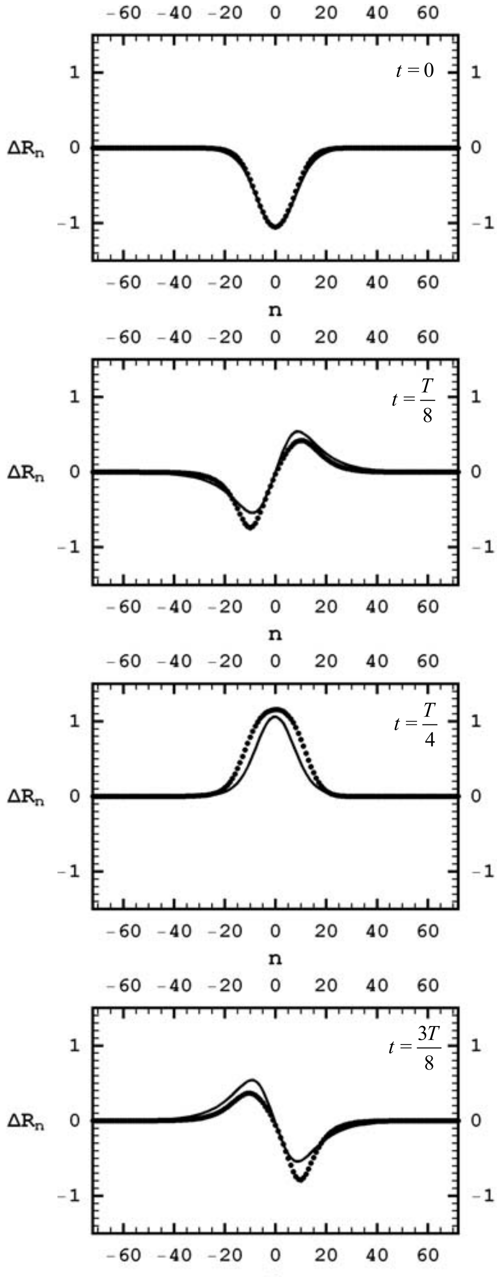



**Fig. 4.** Dots give the breather formed by a chain of stars with initial conditions on the separatrix surrounding the periodic orbit $x_2$. The thin solid line represents the analytic solution in (80).

If the sequence $\{\mathbf{n}\}$ of stars is defined with initial conditions on any invariant curve of Fig.3, then sn-oidal solitons are formed according to the solution (70) travelling along the chain of stars with velocity w. In other words stars pass through a constant phase $\phi$ of the soliton with speed w. This means that a constant phase $\phi$ is stationary with respect to the X-Y frame. This is a very convenient property to define nonlinear density waves or solitary density waves in galaxies.

## 9 Summary and Discussion

We have shown how it is possible to write PDEs governing the motion of collections of non-interacting particles moving in a given potential. In other words, we have shown how one can pass from the theory of orbits to the theory of nonlinear waves.

In particular, we have shown that, on the basis of the third integral, it is possible to write PDEs, governing the motion of stars in galactic models. The fact that the Third Integral can be directly related to the theory of solitons joins different points of view and bridges two different fields of experience.

In the case of the Inner Lindblad resonances a Sine-Gordon PDE is derived for the slow angle of precession of the orbits. The corresponding Lagrangian or Hamiltonian densities are given.

Using the successive consequents of an orbit on the PSOS, we can define a Klein-Gordon chain of stars, that is, an one-dimensional grid obeying a Frenkel-Kontorova Hamiltonian.

Breathers can be constructed in galactic models for sets of stars on the separatrices of unstable periodic orbits. An analytic breather solution is obtained near the ILR which is in good agreement with the corresponding numerical results.

This analysis opens the possibility that collections of orbits forming solitons and breathers can be superimposed, instead of single orbits, to construct bars or spiral arms in galaxies. This approach takes into account the fact that many stars in a galaxy can move in phase correlated orbits, consistent to the structure of phase space. Such collections of stars resist to dispersion as long as the Third Integral is conserved. This feature is important for nonlinear density waves, which can form longer living patterns. The surface of section is only used as a tool to realize the phase correlations of orbits. Notice, that the distribution of stars along an invariant curve can be far from uniform, seriously deformed by the nonlinear effects. This is taken into account as well. In this aspect, the solutions of solitons and breathers, we have found, are more natural building blocks of nonlinear density waves.

**Acknowledgments**

I wish to thank Prof. G. Contopoulos and Dr. Ch. Efthymiopoulos for their comments and helpful discussions. I also wish to thank C. Kalapotharakos for his help in preparing the figures of this paper.

## References


1. Abramowitz M. and Stegun I.A.: "Handbook of Mathematical Functions" (Dover Publications, New York, 1972)
2. Aubry S.: Physica D, **71**, 196 (1994)
3. Bountis T., Capel H.W., Kollemann J.C., Bergamin, J.M. and van der Weele J.P.: Phys. Let. A, **268**, 50 (2000)
4. Bountis T., Bergamin J.M. and Basios V.: Phys. Let. A, **295**, 115 (2002)
5. Contopoulos G.: Zs. f. Ap., **49**, 273 (1960)
6. Contopoulos G.: ApJ, **201**, 566 (1975)
7. Contopoulos G. and Papayannopoulos Th.: A&A, **92**, 33 (1980)
8. Elmegreen B.G. and Thomason M.: A&A, **272**, 37 (1993)
9. Flash S. and Willis C.R.: Phys. Rep., **295**, 181 (1998)
10. Frenkel Ya. and Kontorova T.: Fiz. Zh., **1**, 137 (1939)
11. Kalnais A.: ApJ., **166**, 275 (1971)
33. Kevrekidis P.G., Saxena A. and Bishop A.R.: Phys. Rev. E, **64**, 026611 (2001)
13. Lynden-Bell D. in "Dynamical Structrure and Evolution of Stellar Systems", Saas Fee (Geneva Observatory, 1973)
14. Lynden-Bell D., in "Galactic Dynamics and N-body Simulations", G. Contopoulos, N.K. Spyrou, L. Vlahos (Eds), EADN Astroph. School VI, (Springer-Verlag, 1994)
15. MacKay R.S. and Aubry S.: Nonlinearity, **7**, 1623 (1994)
16. Palmer, P.L.: "Stability of Collisionless Stellar Systems" (Kluwer Academic Publishers, 1994)
17. Patsis P.A., Kaufmann D.E.: A&A, **352**, 469 (1999)
18. Takeno S., Kisoda K. and Sievers A. J., 1988, Prog. Theor. Phys. Suppl. **94**, 242 (1999)
19. Thomason M., Elmegreen, B.G., Donner K.J. and Sundelius B.: ApJ, **356**, L9 (1990)
20. Voglis N., Kalapotharakos C. and Stavropoulos I., 2002, MNRAS, **337**, 619 (2002)
21. Volgis N., Kalapotharakos C., Stavropoulos I. and Efthymiopoulos Ch.: in "Galaxies and Chaos" G. Contopoulos, N. Voglis (eds) Lecture Notes in Physics (Springer-Verlag, 2002)
22. Sievers A.J. and Takeno S.: Phys. Rev. Lett., **61**, 970 (1988)
23. Tsironis G.P.: J. Phys. A: Math. Gen. **35**, 951 (2002)

# Discrete Breathers in Nonlinear Lattices: A Review and Recent Results

Tassos Bountis and Jeroen M. Bergamin

Department of Mathematics and Center for Research and Applications of Nonlinear Systems, University of Patras, 26500 Patras, Greece

**Abstract.** Localization phenomena in systems of many (often infinite) degrees of freedom have attracted attention in solid state physics, nonlinear optics, superconductivity and quantum mechanics. The type of localization we are concerned with here is *dynamic* and refers to oscillations occurring not because of the presence of some defect, but due to the interaction between nonlinearity and resonances. In particular, we shall describe an entity called *discrete breathers*, which represent localized periodic oscillations in nonlinear lattices. As suggested by other authors in this volume, this type of behavior may be observed in density fluctuations of stars rotating in a galaxy in the discrete or continuum approximation. Since the reader may not be too familiar with these concepts, we have chosen first to review the history of discrete breathers in the second half of last century and then present an account of our recent results on the efficient computation of breathers in multi-dimensional lattices using homoclinic orbits. This allows us to make a much more detailed study and classification of discrete breathers than had previously been possible, as well as accurately follow their existence and stability properties as certain physical parameters of the problem are varied.

## 1 The History of Energy Localization

In 1955, Fermi, Pasta and Ulam (FPU) [1] presented the first systematic study of the energy properties of a chain of 32 identical particles, interacting through linear and nonlinear forces and attached to fixed boundaries, according to the equations of motion:

$$\begin{aligned} \ddot{u}_n &= (u_{n+1} - 2u_n + u_{n-1}) + \alpha\left((u_{n+1} - u_n)^2 - (u_n - u_{n-1})^2\right) , \quad n = 1, ..., 31 \\ \dot{u}_0 &= \dot{u}_{32} = 0 \\ u_0 &= u_{32} = 0 , \end{aligned} \tag{1}$$

where $u_n = u_n(t)$ represents each particle's displacement from equilibrium and dots denote differentiation with respect to time t. Using the newly developed computers of the Los Alamos Laboratories in the USA, they integrated (1), starting with the initial condition $u_n = \sin\left(\frac{\pi n}{32}\right)$, and observed a remarkable near-recurrence of the solutions to the initial condition after relatively short time periods (see Fig. 1 below).

All expectations of the theory of statistical mechanics before these experiments predicted that higher order modes of oscillation (i.e. states with $u_n = \sin\left((2k+1)\frac{\pi n}{32}\right)$, $k \in \mathbb{N}$ larger than one) would equally share the energy of

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 75–90, 2003.

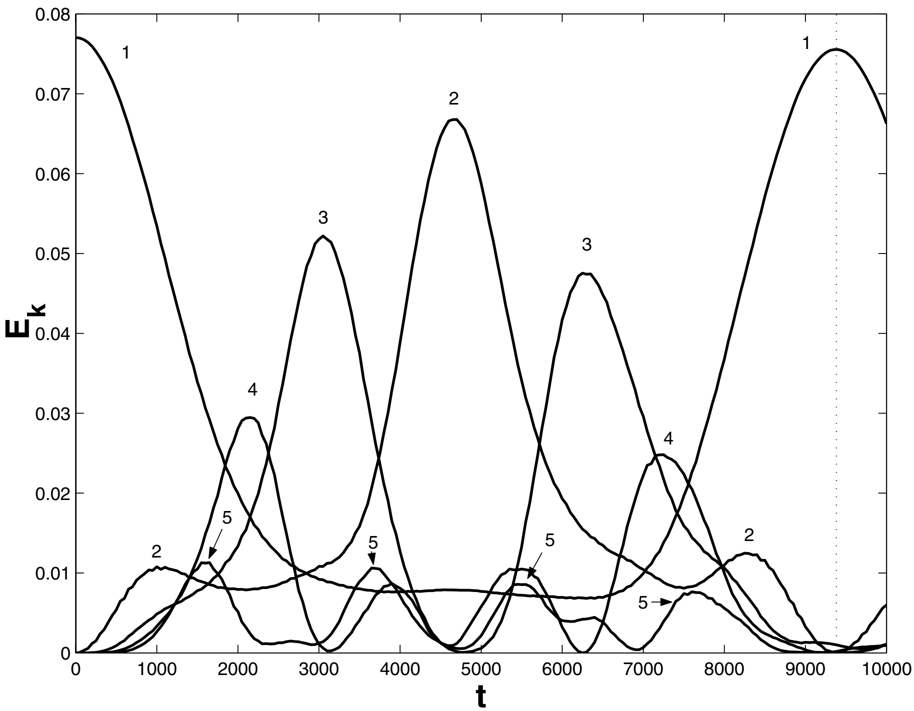



**Fig. 1.** Time-integration of the FPU system until just after the first approximate recurrence of the initial state. Shown here is how the energy $E_k = \frac{1}{2}\left(\dot{a}_k^2 + 2a_k^2 \sin^2\left(\frac{k\pi}{2N}\right)\right)$ is divided over the first five modes $a_k = \sum_{i=1}^{N} x_i \sin\left(\frac{ik\pi}{N}\right)$. The numbers in the figure indicate the wave-number $k$

the system, thus achieving finally a situation of thermodynamic equilibrium. Of course, the Poincaré recurrence theorem dictated that the initial state would again emerge, but after much longer times than the recurrences observed by Fermi, Pasta and Ulam. It was therefore understandable that this discovery created a great excitement within the scientific community of the period.

In 1965, attempting to explain the results of the FPU experiment, Zabusky and Kruskal [2] derived the Korteweg–De Vries (KdV) equation of shallow water waves in the long wavelength and small amplitude approximation,

$$q_\tau + qq_x + \delta^2 q_{xxx} = 0 \quad \delta^2 \ll 1 \; , \tag{2}$$

as a continuum limit of the FPU system (1). They observed that solitary traveling wave solutions, now known as *solitons*, exist, whose interaction properties result in analogous recurrences of initial states, see Fig. 2. In taking their continuum limit, however, Zabusky and Kruskal overlooked an important aspect of the FPU chain: the *discrete nature* of the system.

In 1969, Ovchinnikov [3] showed – in a study of coupled nonlinear oscillators modeling finite-sized molecules – how discreteness in combination with an

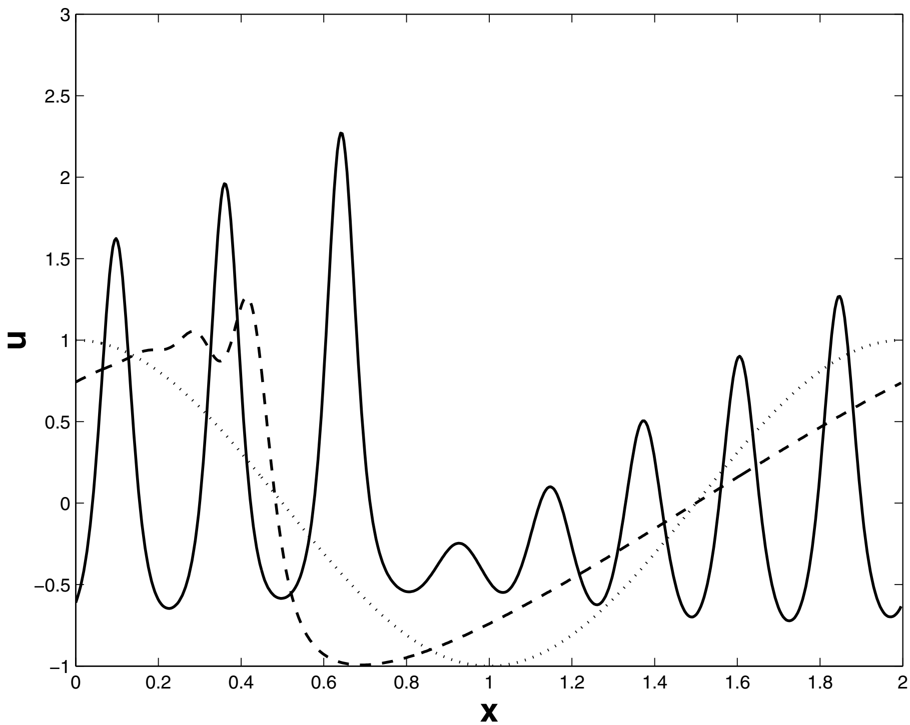



**Fig. 2.** The evolution of the initial state $u = \cos(\pi x)$ in the Korteweg–de Vries equation, at the times $t = 0$ (*dotted line*) $t = 1/\pi$ (*dashed line*). At an intermediate stage ($t = 3.6/\pi$, *solid line*), the solitons are maximally separated, while at $t \approx 30.4/\pi$ the initial state is nearly recovered

intrinsic nonlinearity of the system can cause energy localization. Due to this combination, resonances between neighboring oscillators are avoided when certain frequency bands (the so-called *phonon bands*) are outside the spectrum of vibrations of the system (see Fig. 3). Thus, the recurrence phenomena in the FPU experiments can be explained as the result of limited energy transport between Fourier modes, caused by discreteness and non-resonance effects.

### 1.1 The Discovery of Discrete Breathers

Twenty years later (1988), the subject of localized oscillations in nonlinear lattices was revived in a paper by Sievers and Takeno [4], in which they used analytical arguments to show that energy localization occurs generically in FPU systems of *infinitely many* particles in one dimension, obeying the equations:

$$\ddot{u}_n = (u_{n+1} - 2u_n + u_{n-1}) + \alpha \left( (u_{n+1} - u_n)^2 - (u_n - u_{n-1})^2 \right) , \qquad (3)$$

cf. (1), for $-\infty < n < \infty$.

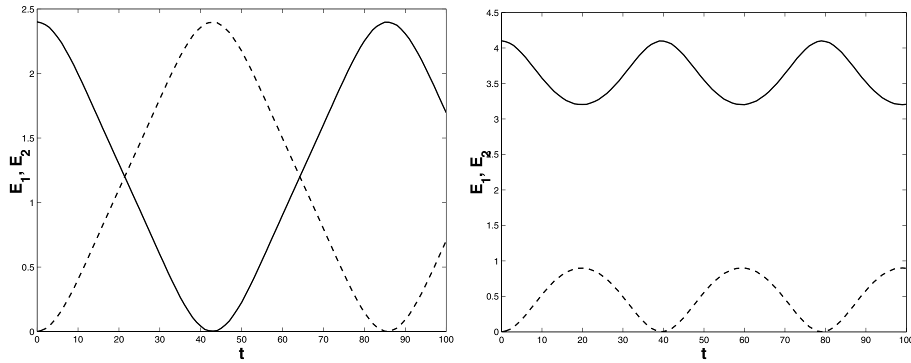



**Fig. 3.** Local energies $E_1(t)$ (*solid line*) and $E_2(t)$ (*dashed line*) of the two coupled nonlinear oscillators considered by Ovchinnikov [3], showing how energy transfer is impeded by discreteness and nonlinearity. **Left:** Complete energy transfer when the first oscillator has initial amplitude $a = 2.0$ ($E_1(0) = 2.4$). **Right:** Incomplete energy transfer when the first oscillator has initial amplitude $a = 2.5$ ($E_1(0) \approx 4.1$)

Combining their perturbative analysis with numerical experiments, they demonstrated that a new type of solution, the so-called *discrete breathers* exist as oscillations which are both time-periodic and spatially localized. In their simplest form, such solutions can exhibit significant oscillations only of the middle ($n = 0$) and nearby ($n = -1, +1$) particles. However, a great many patterns are possible (the so-called *multibreathers*) in which several particles around the middle one oscillate with large amplitudes, as shown here in Fig. 4. How can one determine all the possible shapes? Which of them are stable under small perturbations? These are the kind of questions that we set out to answer in our research.

Besides the FPU system, the existence of these localized oscillations was soon verified numerically by other research groups on a variety of lattices, including the Klein-Gordon (KG) system

$$\ddot{u}_n = -\,V'(u_n) + \alpha\,(u_{n+1} - 2u_n + u_{n-1}) \ , \tag{4}$$

where $V$ is an on-site potential and $\alpha$ is a parameter indicating the coupling strength.

It was not, however, until 1994, that a mathematical proof of the existence of discrete breathers was published by MacKay and Aubry [5] in the case of one-dimensional lattices of the type (4). Under the general assumptions of nonlinearity and non-resonance, such chains of interacting oscillators were rigorously shown to possess discrete breather solutions for small enough values of the coupling parameter $\alpha > 0$, as a continuation of their obvious existence at $\alpha = 0$.

Section 2 below contains the main part of our contribution to the field of discrete breathers. We have found that a very convenient way to construct them and study their stability properties, away from the $\alpha = 0$ limit, is through the "*geometry*" of the *homoclinic solutions* of *nonlinear recurrence relations* [6].

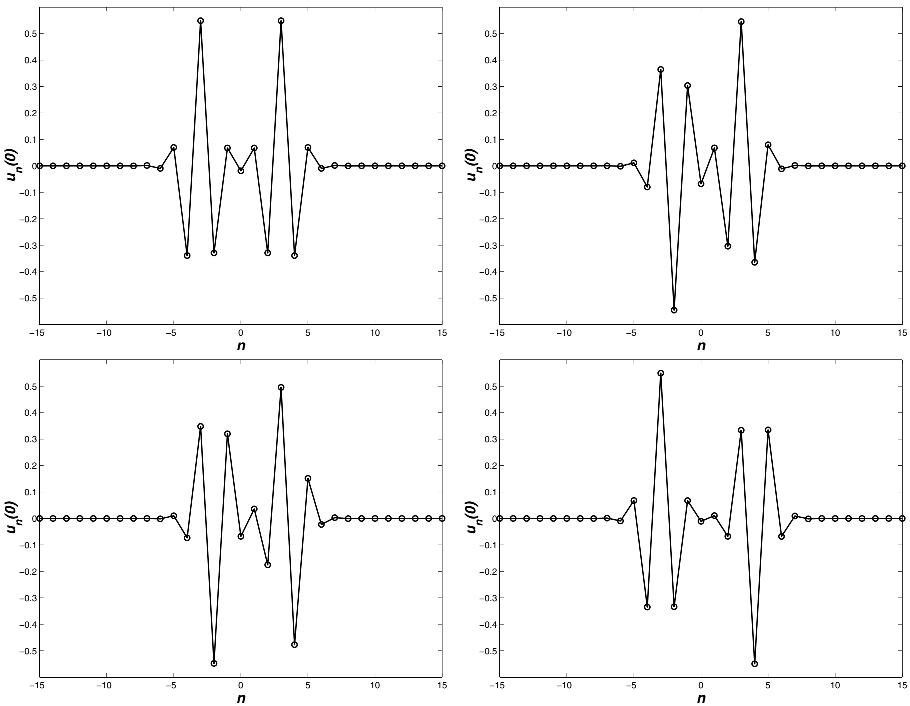



**Fig. 4.** Several breathers of an FPU lattice of the form (3) with cubic (rather than quadratic) interactions, obtained by starting a Newton-Raphson search using homoclinic orbits of a map of the form (6) as an initial guess

## 2 The Connection with Homoclinic Dynamics

Focusing on the property of spatial localization, Flach was the first to show that discrete breathers in simple one-dimensional chains can be actually represented by homoclinic orbits in the Fourier amplitude space of time-periodic functions [7]. Indeed, inserting a Fourier series

$$u_n(t) = \sum_{k=\infty}^{\infty} A_n(k) \exp(\mathrm{i}k\omega t) \tag{5}$$

into the equations of motion of either the FPU (3) or KG (4) lattice and setting the amplitudes of terms with the same frequency equal to zero, leads to the system of equations

$$\begin{aligned} -k^2\omega^2 A_n(k) = \langle -V'(u_n) + W'(u_{n+1} - u_n) - W'(u_n - u_{n-1}), \exp(\mathrm{i}k\omega t)\rangle \\ \forall k, n \in \mathbb{Z}\,. \end{aligned}$$

This is an infinite-dimensional mapping of the Fourier coefficients $A_n(k)$ with the brackets $\langle .,. \rangle$ indicating a properly normalized inner product. Time-periodicity is ensured by the Fourier basis functions $\exp(\mathrm{i}k\omega t)$. Spatial localization requires that $A_n(k) \to 0$ exponentially as $n \to \pm\infty$. Hence a discrete breather is a homoclinic orbit in the space of Fourier coefficients, i.e. a doubly infinite sequence of points beginning at 0 for $n \to -\infty$ and ending at 0 for $n \to +\infty$. In fact, keeping only the Fourier term ($k = 1$) with the largest amplitude reduces, in some cases, the above system to a simple 2-dimensional mapping

$$a_{n+1} = g(a_n, a_{n-1}) \ , \tag{6}$$

whose invariant manifolds of the saddle fixed point at the origin can be easily plotted, as in Fig. 5.

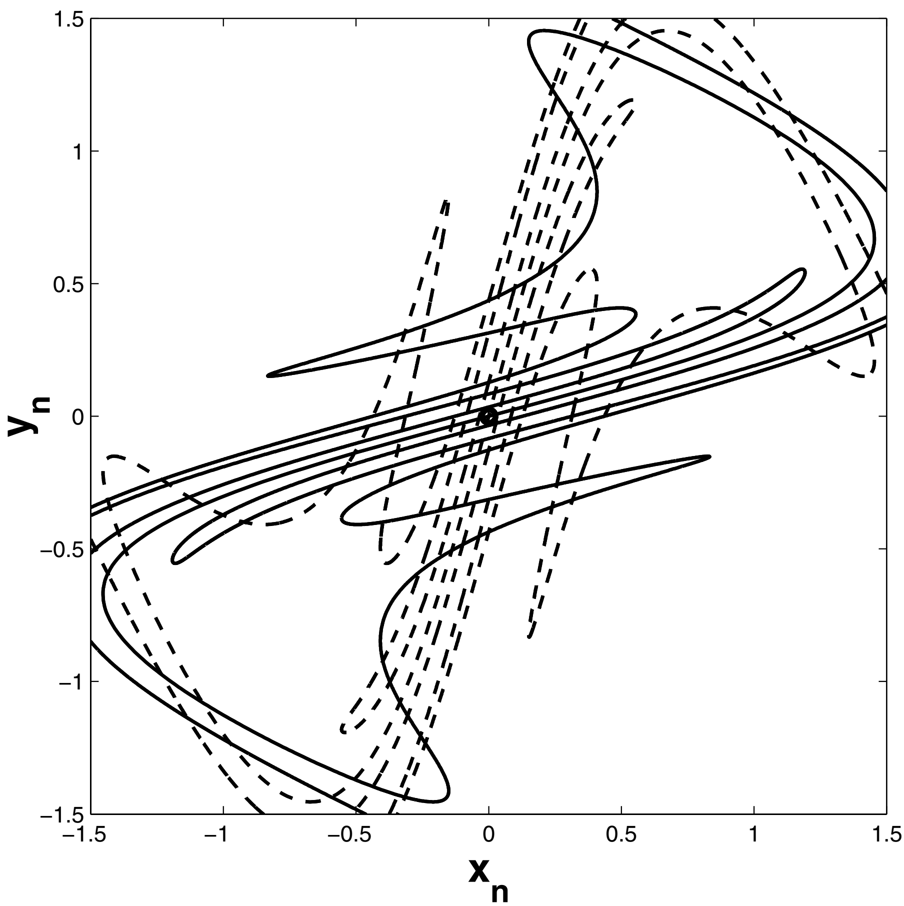



**Fig. 5.** The stable (*dashed line*) and unstable (*solid line*) manifolds of the fixed point at $(0, 0)$ of a 2-dimensional map of the form (6), with $x_n = a_n$ and $y_n = a_{n+1}$. The manifolds are clearly seen to intersect at infinitely many points, hence a wealth of homoclinic orbits exists

### 2.1 How to Construct Homoclinic Orbits

Realizing the importance of homoclinic orbits to the subject of discrete breathers, we have been able to develop in [8] efficient numerical methods for locating homoclinic orbits of invertible maps of arbitrary but finite dimension. To this end, we found it particularly useful to exploit symmetry properties of the maps and understand the geometry of the invariant manifolds near the origin, which is always a fixed point of the mappings of the saddle type.

To see how this is done, let us consider a general first order map (or recurrence relation)

$$\boldsymbol{x}_n = f\left(\boldsymbol{x}_n\right) \ , \quad \boldsymbol{x}_n \in \mathbb{R}^d \ , \tag{7}$$

with $d$ a positive integer. Recall that homoclinic orbits are solutions for which $\boldsymbol{x}_n \to 0$ as $n \to \pm\infty$ and concentrate on all orbits satisfying a symmetry condition of the form

$$\boldsymbol{x}_n = \mathrm{M}\boldsymbol{x}_{-n} \ , \tag{8}$$

where M is a $d \times d$-matrix with constant entries and $\det\left(\mathrm{M}\right) \neq 0$. Observe that if an orbit obeys such a symmetry and $\boldsymbol{x}_n \to 0$ as $n \to -\infty$, then it is also true that $\boldsymbol{x}_n \to 0$ as $n \to \infty$. Hence, any orbit which obeys this symmetry and satisfies $\boldsymbol{x}_n \to 0$ as $n \to -\infty$ is a homoclinic orbit.

To obtain such an orbit, we first need to specify numerically its asymptotic behavior as $n \to -\infty$. In other words, it has to be on the *unstable manifold* of $\boldsymbol{x} = 0$. Furthermore, a necessary requirement for a homoclinic orbit to exist is that the origin be a saddle fixed point of the map. Then it is known that the unstable manifold of the origin is well approximated in its vicinity by the unstable Euclidean eigenspace of the linearized equations, which is tangent to the nonlinear manifold and has the same number of dimensions.

Now, the dimension of the linear unstable eigenspace equals the number of coordinates necessary to determine a point $\boldsymbol{x}_{-N}$, $N \gg 1$ uniquely. Thus, by choosing this point to be on the linear unstable manifold very close to the origin, it follows that it will also be approximately on the corresponding nonlinear manifold. Thus, when mapped forward $N+1$ times, we can test whether it satisfies the above symmetry relation (8). By this approach, locating symmetric homoclinic orbits becomes a search for solutions of the system

$$\begin{cases} \boldsymbol{x}_1 - \mathrm{M}\boldsymbol{x}_{-1} = 0 \\ \boldsymbol{x}_0 - \mathrm{M}\boldsymbol{x}_0 = 0 \end{cases} , \tag{9}$$

since, given $\boldsymbol{x}_{-N}$, the values of $\boldsymbol{x}_1$, $\boldsymbol{x}_0$ and $\boldsymbol{x}_{-1}$ are uniquely obtained by direct iteration of the map (7).

In the case of an invertible map we can use this method to find also all *asymmetric* homoclinic orbits, i.e. those which do not obey the symmetry condition (9). This can be done by introducing the new "sum" and "difference" variables

$$\begin{cases} \boldsymbol{v}_n = \boldsymbol{x}_n + \boldsymbol{x}_{-n} \\ \boldsymbol{w}_n = \boldsymbol{x}_n - \boldsymbol{x}_{-n} \end{cases} ,$$

which always possess the symmetry

$$\begin{cases} \boldsymbol{v}_n = \boldsymbol{v}_{-n} \\ \boldsymbol{w}_n = -\boldsymbol{w}_{-n} \end{cases} . \tag{10}$$

In this way, we can apply again the above strategy and look for symmetric homoclinic orbits of a new map (of *double the dimension* of the original $f$) described by the equations

$$F : \begin{cases} \boldsymbol{v}_{n+1} = f\left(\frac{\boldsymbol{v}_n+\boldsymbol{w}_n}{2}\right) + f^{-1}\left(\frac{\boldsymbol{v}_n-\boldsymbol{w}_n}{2}\right) \\ \boldsymbol{w}_{n+1} = f\left(\frac{\boldsymbol{v}_n+\boldsymbol{w}_n}{2}\right) - f^{-1}\left(\frac{\boldsymbol{v}_n-\boldsymbol{w}_n}{2}\right) \end{cases} , \tag{11}$$

yielding homoclinic orbits $\boldsymbol{x}_n$ of the original map $f$, (7), that are not themselves necessarily symmetric. On the other hand, each homoclinic orbit of $f$ is a symmetric homoclinic orbit of the new map $F$. Therefore, by determining all *symmetric* homoclinic orbits of $F$, we find *all* homoclinic orbits of $f$. Following this approach, we have also been able to *classify* all possible homoclinic orbits by assigning to them *symbolic sequences* in a systematic way, according to their complexity (see [8, 9] for more details and Fig. 6 as an example of the results of these papers).

Thus, we now come to our second major contribution on this topic, described in Sect. 3 below. This concerns a new approach to the computation of discrete breathers, which can be efficiently applied to lattices of more than one spatial dimension and systems with vector valued variables assigned to each lattice site. The main idea is to write a breather solution as a product of a space-dependent and a time-dependent part and reduce the problem to finding the homoclinic orbits of a 2 dimensional map, under the constraint that the given ODEs possess simple periodic oscillations of a well-defined type and of specified period.

## 3 A New Approach Is Introduced

As was mentioned above, it is possible to write down a map in Fourier amplitude space linking discrete breathers with homoclinic orbits. This map can be reduced to a finite-dimensional recurrence relation, by neglecting Fourier components with a wave number $k$ larger than some cutoff value $k_{max}$. Then, one can use the methods of our papers [8, 9] to approximate discrete breather solutions by finding all the homoclinic solutions of these recurrence relations.

Recently, however, this problem has been re-examined from a different perspective: In 2002, inspired by the work of other authors like Flach [10] and Kivshar [11], Tsironis [12] suggested a new way to approximately *separate* amplitude from time-dependence, yielding in some cases ODEs with known solutions (for example elliptic functions) while keeping the dimension of the recurrence relation as low as possible. This led to an improved accuracy of the calculations and provided analytical expressions of discrete breathers for a special class of FPU and KG systems.

In a very recent paper [13], Bergamin extended Tsironis' work by developing a numerical procedure for which the time-dependent functions need not be

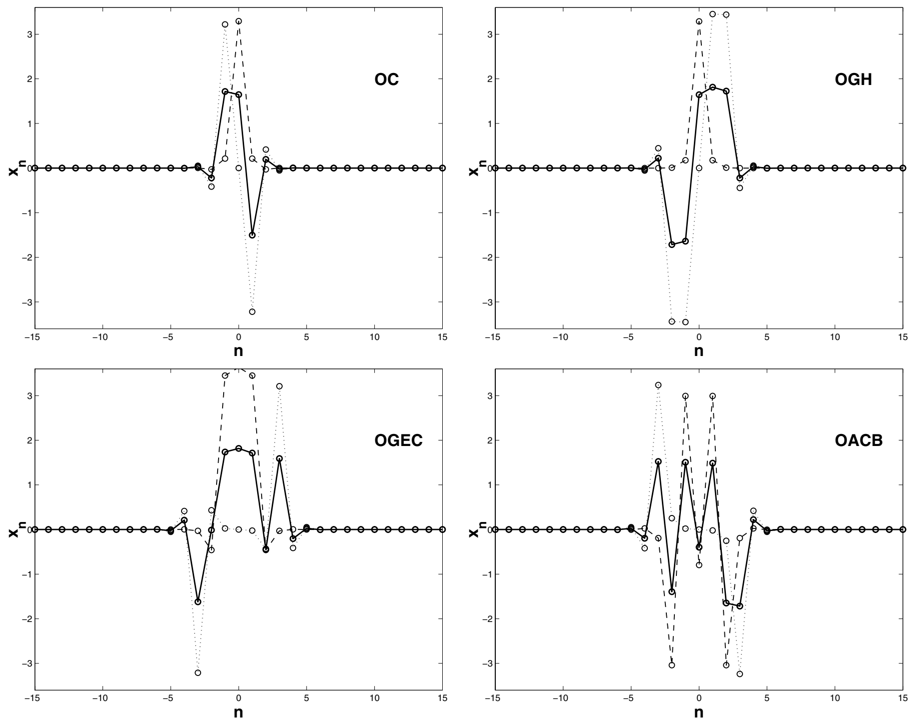



**Fig. 6.** Several homoclinic orbits of the map $F$, (11) determined by a zero-search of the system of equations $v_1 - v_{-1} = 0$ and $w_0 = 0$, related by $v_n = x_n + x_{-n}$ and $w_n = x_n - x_{-n}$. Shown here are $v_n$ (*dashed*), $w_n$ (*dotted*) and $x_n$ (*solid*). Also indicated is the symbolic name of the orbit, assigned by the procedure given in [8]

known analytically. In this way, a much wider class of nonlinear lattices can now be treated involving scalar or vector valued variables in one or more spatial dimensions.

In particular, the approximation proposed by Tsironis can be more precisely formulated as follows

$$\begin{cases} u_{n+1}(t) - u_n(t) \approx (a_{n+1} - a_n) T_n(t) \\ u_{n-1}(t) - u_n(t) \approx (a_{n-1} - a_n) T_n(t) \end{cases}, \tag{12}$$

where $a_n$ denotes the time-independent amplitude of $u_n(t)$ and $T_n(t)$ is its time-dependence, defined by $T_n(0) = 1$ and $\dot{T}_n(0) = 0$.

Note now, that all FPU and KG systems are derived from a potential function of the form

$$U = \sum_{n=-\infty}^{\infty} V(u_n) + W(u_{n+1} - u_n) \ .$$

Using the approximation (12), the equations of motion

$$\ddot{u}_n = -V'(u_n) + W'(u_{n+1} - u_n) - W'(u_n - u_{n-1})$$

are transformed into

$$a_n \ddot{T}_n = -V'(a_n T_n) + W'((a_{n+1} - a_n) T_n) - W'((a_n - a_{n-1}) T_n) \ . \quad (13)$$

This is an ordinary differential equation (ODE) for $T_n(t)$ which can, in principle, be solved since the initial conditions are known.

Of course, an analytical solution of ODE (13) is in general very difficult to obtain. However, since we are primarily interested in the amplitudes $a_n$, what we ultimately need to do is develop a numerical procedure to find a recurrence relation linking $a_n$, $a_{n+1}$ and $a_{n-1}$ *without having to solve the ODE beforehand.*

Let us observe first, that the knowledge of $a_n$, $a_{n+1}$ and $a_{n-1}$ permits us to solve the above ODE numerically. Under mild conditions, solutions $T_n(t)$ can thus be obtained, which are time-periodic, while for a discrete breather all functions $T_n(t)$ have the same period. Choosing a specific value for this period, allows us to *invert the process* and determine $a_{n+1}$ as a function of $a_n$ and $a_{n-1}$, similar to (6). In the same way, we also determine $a_{n-1}$ as a function of $a_n$ and $a_{n+1}$. Thus, a *two-dimensional invertible map* for the $a_n$ has been constructed, ensuring that all oscillators have the same frequency.

As is explicitly shown in [13, 14], on a variety of examples, the homoclinic orbits of this map provide highly accurate approximations to the discrete breather solutions with the given period and the initial state $u_n(0) = a_n$, $\dot{u}_n(0) = 0$. In fact, we can now apply this approach to more complicated potentials and higher dimensional lattices, as we demostrate in Fig. 7, where we compute a discrete breather solution of a 2-dimensional lattice, with indices $n, m$ in the $x, y$ directions and dependent variable $u_{n,m}(t)$.

Having thus discovered new and efficient ways of calculating discrete breathers in a wide class of nonlinear lattices, we now turn to the study of their stability, control and continuation properties in parameter space. More specifically, we shall show that it is possible to use our methods to extend the domain of existence of breathers to parameter ranges that cannot easily reached by other more standard continuation techniques.

### 3.1 Stability and Existence of Discrete Breathers Using Control

So far, we have seen that transforming nonlinear lattice equations to low-dimensional maps and using numerical methods to compute their homoclinic orbits provides an efficient tool for approximating discrete breathers in any (finite) dimension and classifying them in a systematic way. This clears the path for an investigation of important properties of large numbers of discrete breathers of increasing complexity. One such property, which is relevant to many applications and requires that a solution be known to great accuracy, is stability in time.

In our recent work [15], the accurate knowledge of a discrete breather solution was used in a rather uncommon way to study stability properties: As is well known, in Physics, Electronics and Engineering, a familiar task is to try to influence the behavior of a system, by applying control methods. By control we mean here the addition of an *external force* to the system which allows us

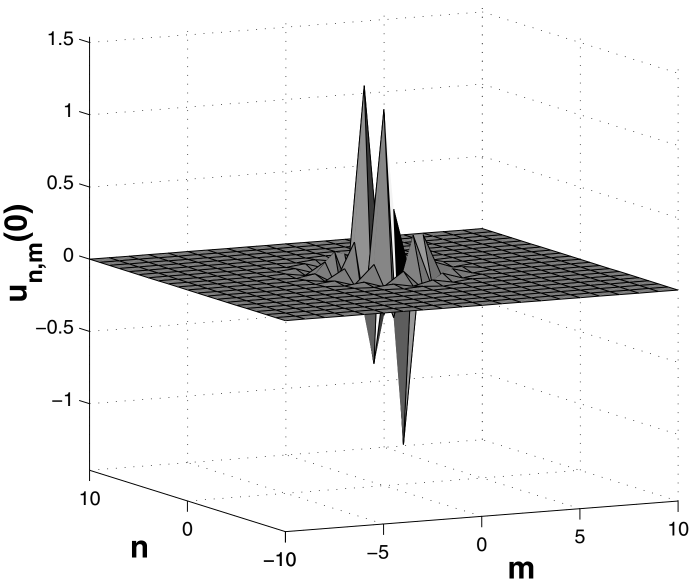



**Fig. 7.** A 2-dimensional breather obtained by the method described in the text, for equations of motion of the KG type where the particles at each lattice site $n, m$ experience harmonic interactions with its 4 nearest neighbors and a quartic on-site potential.

to influence its dynamics. In particular, the objective of our control will be to change the *stability type* of a discrete breather solution of the controlled system, compared with the uncontrolled one.

The system is altered in such a way that the solution itself does not change. In other words, in the controlled system the solution is exactly the same as in the uncontrolled and for this reason the control is of the feedback type. If a solution is unstable in the uncontrolled system, the extra terms added to the equations in the controlled case can cause the solution to become stable and vice versa. The system we have studied [15] is based on the KG equations of motion written as

$$\ddot{u}_n = -V'(u_n) + \alpha(u_{n+1} - 2u_n + u_{n-1}) + L\frac{d}{dt}(\hat{u}_n - u_n) \ , \tag{14}$$

where $\hat{u}_n$ is the known discrete breather solution of the equations when $L = 0$. The parameter $L$ indicates how strongly the control term influences the KG system. Thus, for any value of $L$, $u_n = \hat{u}_n$ is clearly seen to be a solution of both the controlled as well as the uncontrolled lattice equations. Clearly, for $L > 0$, the $L\frac{d}{dt}u_n$ term introduces dissipation, while the $L\frac{d}{dt}\hat{u}_n$ represents periodic forcing. It is therefore reasonable to investigate whether, by increasing $L$, the dissipative part of the process will force the system to converge in time to a stable solution. If this solution is the original $\hat{u}_n$, the latter is stable. If this does not happen, the original solution is unstable. In [14, 17] the following proposition is proved, though in a slightly more general formulation:

**Proposition 1.** *Let $u_n = \hat{u}_n$ be a periodic solution of the lattice equations*

$$\ddot{u}_n = -V'(u_n) + \alpha(u_{n+1} - 2u_n + u_{n-1}) \ .$$

*Then there exists an $L > 0$ such that $u_n = \hat{u}_n$ is an asymptotically stable solution of the modified (controlled) system*

$$\ddot{u}_n = -V'(u_n) + \alpha(u_{n+1} - 2u_n + u_{n-1}) + L\frac{d}{dt}(\hat{u}_n - u_n) \ .$$

This result clearly implies that, by increasing $L$, it is possible to stabilize the original solution, independent of its stability in the uncontrolled situation. Let us demonstrate this by taking the breather of Fig. 8, which is unstable at $L = 0$, substitute its (known) form $\hat{u}_n(t)$ in the above equations and increase the value of $L$. As we see in Fig. 5, it is quite easy to stabilize it at $L = 1.17$, since increasing the value of $L$ gradually brings all eigenvalues of the monodromy matrix of the solution inside the unit circle.

The above Proposition has an additional significant advantage: It gives us the opportunity to address the question of the *existence* of discrete breathers in ranges of the coupling parameter $\alpha$ where other techniques do not apply. In order to do this, it is important to recall first how this question was originally answered (though partially) in the existence proof of MacKay and Aubry, based on the notion of the so - called anti-continuum limit $\alpha = 0$ [5].

Let us observe that the KG equations of motion

$$\ddot{x}_i = -\,V'(x_i) + \alpha(x_{i+1} - 2x_i + x_{i-1})$$

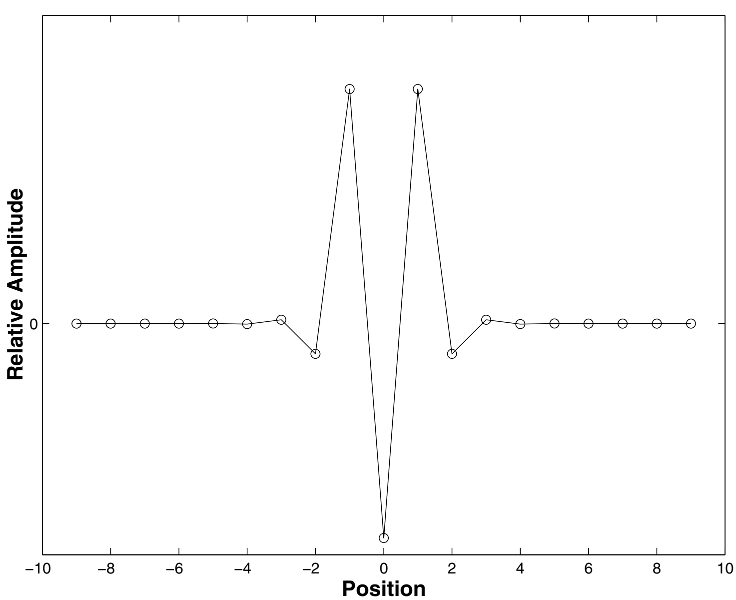



**Fig. 8.** A breather shape $u_n(t) = \hat{u}_n(t)$ of the system (14) which is unstable for $L = 0$

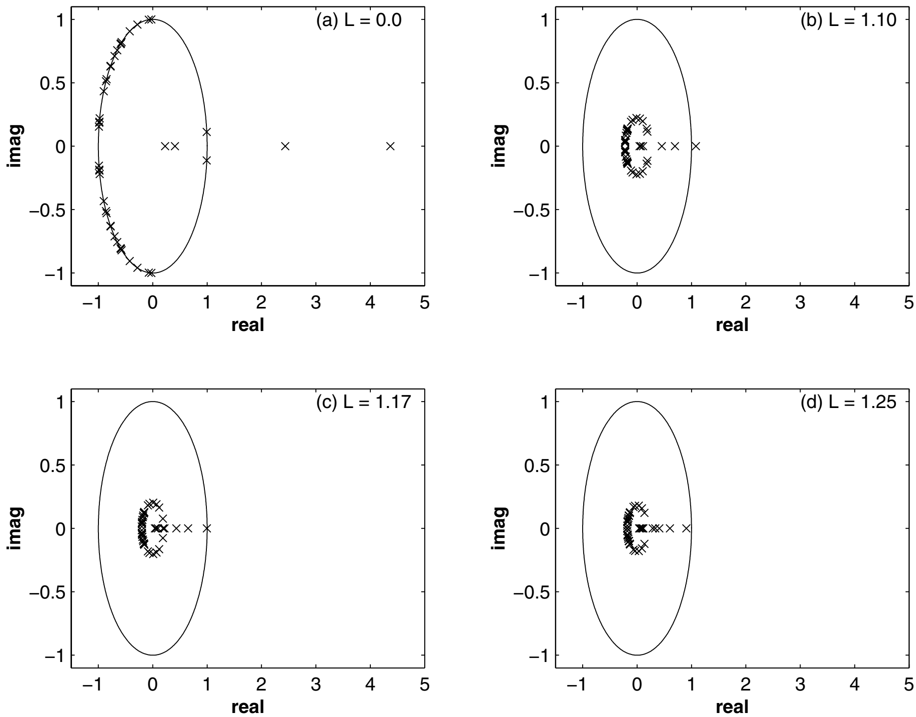



**Fig. 9.** Increasing the control parameter $L$ from $L = 0$ in the system (14) with the initial shape given in Fig. 8 moves the eigenvalues of the orbit's monodromy matrix inside the unit circle. Initially, some eigenvalues are outside the unit circle, but eventually, for $L > 1.17$, all eigenvalues attain magnitudes less than one, thus achieving stability and control

describe a system of uncoupled oscillators for $\alpha = 0$. Obviously, in that case, any initial condition, where only a finite number of oscillators have a non-zero amplitude, is a discrete breather solution. In their celebrated paper of 1994, MacKay and Aubry prove that, under the conditions of nonresonance with the phonon band and nonlinearity of the function $V'(x)$, this solution can be continued to the regime where $\alpha > 0$.

According to their approach, however, continuation for $\alpha > 0$ is possible, only as long as the eigenvalues of the Floquet matrix of the solution do not cross the value $+1$. This means that the typical occurrence of a bifurcation, through which the breather becomes unstable, prevents MacKay and Aubry's continuation method from following the breather beyond that value of $\alpha$. This is where the control method we have proposed comes to the rescue: As we can see in Fig. 7, by following a path in $\alpha > 0$ and $L > 0$ space, a discrete breather solution can be continued to a higher value of $\alpha$, by choosing $L$ in the controlled system such that the solution remains stable!

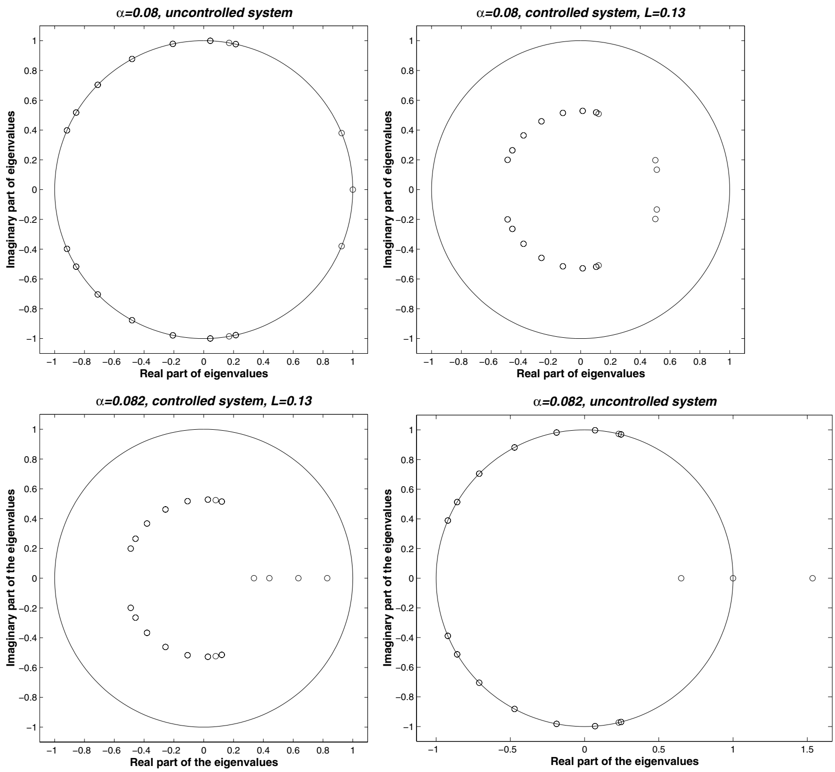



**Fig. 10.** When the uncontrolled system approaches a bifurcation point for $\alpha = 0.08$ and $L = 0$ (**upper left figure**), the bifurcation can be avoided by increasing the control parameter, for example to $L = 0.13$, such that the breather solution is asymptotically stable (**upper right**). When the coupling strength is increased to $\alpha = 0.082$ (**lower left**), control can be switched off to return to the breather of the uncontrolled system, which is now unstable (**lower right**)

Therefore, since by the above Proposition one can always find $L$ such that stability is possible for any coupling $\alpha$, this allows the continuation of any solution of the $\alpha = 0$ case to $\alpha > 0$ values beyond bifurcation, demonstrating the existence of breather solutions in the corresponding parameter regime. The fact that $\hat{u}_n$, which is a (stable) solution of the controlled system is by definition also a (unstable) solution of the uncontrolled system, implies that we have succeeded in continuing a discrete breather solution to higher values of the coupling parameter.

In Fig. 6, we show in $L, \alpha$ space the regions of stability of this particular breather. As is well-known by the work of Segur and Kruskal [16], breathers are not expected to exist in these systems in the continuum limit of $\alpha$ going to infinity. At what coupling value though and how do they disappear? Can we use our control aided continuation methods to follow them at arbitrarily high $\alpha$ to

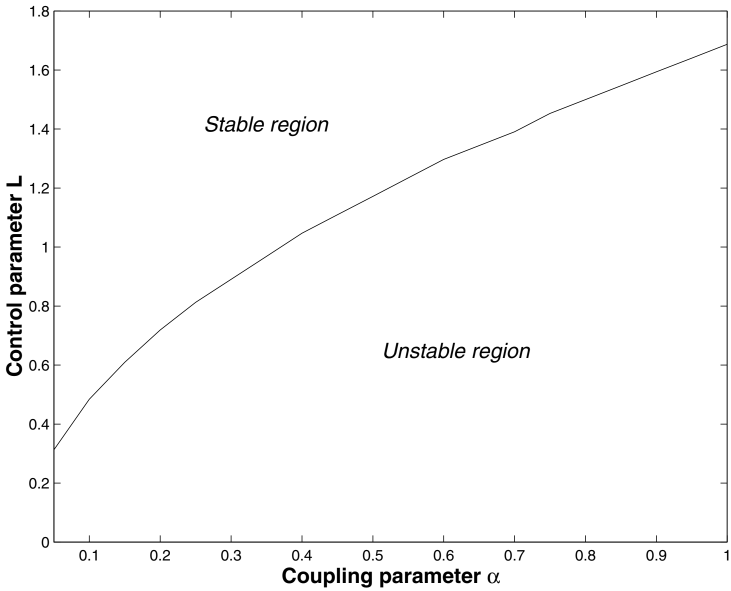



**Fig. 11.** If the coupling $\alpha$ is increased, the control parameter $L$ has to be larger to achieve stability and successful control. Shown here are the regions for which the breather of Fig. 8 is a stable or unstable solution $u_n\,(t) = \hat{u}_n\,(t)$ of (14)

be able to answer such questions? Currently, we are working on this problem and results are expected to appear in a future publication [17].

**Acknowledgements**

One of the authors (T.B.) wishes to thank G. Contopoulos and N. Voglis for their interest in this work and their invitation to present it in a conference on Galaxies and Chaos. The other author (J.M.B.) acknowledges partial support by a "Karatheodory" graduate research fellowship of the University of Patras.

## References

1. E. Fermi, J. Pasta and S. Ulam: Tech. Rep. Los Alamos Nat. Lab. **LA 1940**, (1955)
2. N.J. Zabusky and M.D. Kruskal: Phys. Rev. Lett. **15**, 240 (1965)
3. A.A. Ovchinnikov: Zh. Eksp. Teor. Fiz. **57**, 263 (1969) Sov. Phys. JETP **30**, 147 (1970)
4. A.J. Sievers and S. Takeno: Phys. Rev. Lett. **61**, 970 (1988)
5. R.S. MacKay and S. Aubry: Nonlinearity **7**, 1623 (1994)
6. T. Bountis, H.W. Capel, M. Kollmann, J. Ross, J.M. Bergamin and J.P. van der Weele: Phys. Lett. A **268**, 50 (2000)
7. S. Flach: Phys. Rev. E **51**, 3579 (1995)
8. J.M. Bergamin, T. Bountis and M.N. Vrahatis: Nonlinearity **15**, 1603 (2002)

9. J.M. Bergamin, T. Bountis and C. Jung: J. Phys. A: Math. Gen. **33**, 8059 (2000)
10. S. Flach: Phys. Rev. E **50**, 3134 (1994)
11. Y.S. Kivshar: Phys. Rev. E **48**, R43 (1993)
12. G.P. Tsironis: Journal of Physics A **35**, 951 (2002)
13. J.M. Bergamin: Numerical approximation of breathers in lattices with nearest-neighbor interactions, to appear in Phys. Rev. E (2003)
14. J.M. Bergamin: Localization in nonlinear lattices and homoclinic dynamics. PhD thesis, University of Patras, Greece (2003)
15. T. Bountis, J.M. Bergamin and V. Basios: Phys. Lett. A **295**, 115 (2002)
16. H. Segur and M.D. Kruskal: Phys. Rev. Lett. **58**, 747 (1987)
17. J.M. Bergamin, T. Bountis and H.W. Capel: Continuation of discrete breathers using control methods, in preparation (2003).

# Chaos or Order in Double Barred Galaxies?

Witold Maciejewski

[1] INAF – Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, 50125 Firenze, Italy,
[2] Obserwatorium Astronomiczne Uniwersytetu Jagiellońskiego, Kraków, Poland

**Abstract.** Bars in galaxies are mainly supported by particles trapped around closed periodic orbits. These orbits respond to the bar's forcing frequency only and lack free oscillations. We show that a similar situation takes place in double bars: particles get trapped around orbits which only respond to the forcing from the two bars and lack free oscillations. We find that writing the successive positions of a particle on such an orbit every time the bars align generates a closed curve, which we call a loop. Loops allow us to verify consistency of the potential. As maps of doubly periodic orbits, loops can be used to search the phase-space in double bars in order to determine the fraction occupied by ordered motions.

## 1 Introduction

Bars within bars appear to be a common phenomenon in galaxies. Recent surveys show that up to 30% of early-type barred galaxies contain nested bars [4]. The relative orientation of the two bars is random, therefore it is likely that the bars rotate with different pattern speeds. Inner bars, like large bars, are made of relatively old stellar populations, since they remain distinct in near infrared [5]. Galaxies with two independently rotating bars do not conserve the Jacobi integral, and it is a complex dynamical task to explain how such systems are sustained. To account for their longevity, one has to find sets of particles that support the shape of the potential in which they move. Particle motion in a potential of double bars belongs to the general problem of motion in a pulsating potential [6] [9], of which the restricted elliptical 3-body problem is the best known example. Families of closed periodic orbits have been found in this last problem, where the test particle moves in the potential of a binary star with components on elliptical orbits [2]. However, such families are parameterized by values that also characterize the potential (i.e. ellipticity of the stellar orbit and the mass ratio of the stars), and their orbital periods are commensurate with the pulsation period of the potential. For a given potential, these families are reduced to single orbits separated in phase-space. The solution for double bars is formally identical, and there an orbit can close only when the orbital period is commensurate with the relative period of the bars. Such orbits are separated in phase-space, and therefore families of closed periodic orbits are unlikely to provide orbital support for nested bars. Another difficulty in supporting nested bars is caused by the piling up of resonances created by each bar, which leads to considerable chaotic zones. In order to minimize the number of chaotic zones,

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 91–100, 2003.

resonant coupling between the bars has been proposed [10], so that the resonance generated by one bar overlaps with that caused by the other bar.

Finding support for nested bars has been hampered by the fact that closed periodic orbits are scarce there. However, it is particles, not orbits, which create density distributions that support the potential. The concept of closed periodic orbit is too limiting in investigation of nested bars, and another description of particle motion, which does not have its limitations, is needed. Naturally, in systems with two forcing frequencies, double-periodic orbits play a fundamental role. Thus in double bars a large fraction of particle trajectories gets trapped around a class of double-periodic orbits. Although such orbits do not close in any reference frame, they can be conveniently mapped onto the loops [8], which are an efficient descriptor of orbital structure in a pulsating potential. The loop is a closed curve that is made of particles moving in the potential of a doubly barred galaxy, and which pulsates with the relative period of the bars. Orbital support for nested bars can be provided by placing particles on the loops.

Here I give a systematic description of the loop approach, which recovers families of stable double-periodic orbits, and which can be applied to any pulsating potential. In Sect. 2 I use the epicyclic approximation to introduce the basic concepts, and in Sect. 3 I outline the general method.

## 2 The Epicyclic Solution for Any Number of Bars

If a galaxy has a bar that rotates with a constant pattern speed, it is convenient to study particle orbits in the reference frame rotating with the bar. If two or more bars are present, and each rotates with its own pattern speed, there is no reference frame in which the potential remains unchanged. In order to point out formal similarities in solutions for one and many bars, I solve the linearized equations in the inertial frame. This is equivalent to the solution in any rotating frame, and the transformation is particularly simple: in the rotating frame the centrifugal and Coriolis terms are equivalent to the Doppler shift of the angular velocity. It is convenient to show it in cylindrical coordinates $(R, \varphi, z)$: if $\mathbf{e}_z$ is the rotation axis, then the $R$ and $\varphi$ components of the equation of motion for the rotating frame, $\ddot{\mathbf{r}} = -\nabla\Phi - 2(\mathbf{\Omega_B} \times \dot{\mathbf{r}}) - \mathbf{\Omega_B} \times (\mathbf{\Omega_B} \times \mathbf{r})$, can be written as

$$\ddot{R} - R(\dot{\varphi} + \Omega_B)^2 = -\frac{\partial \Phi}{\partial R},$$
$$R\ddot{\varphi} + 2\dot{R}(\dot{\varphi} + \Omega_B) = -\frac{1}{R}\frac{\partial \Phi}{\partial \varphi}.$$

These equations are identical with the components of the equation of motion in the inertial frame,

$$\ddot{\mathbf{r}} = -\nabla\Phi, \tag{1}$$

where clearly the angular velocity $\dot{\varphi}$ in the rotating frame corresponds to $\dot{\varphi} + \Omega_B$ in the inertial frame. For the rest of this section I assume the inertial frame, in which the equation of motion (1) has the following $R$ and $\varphi$ components in cylindrical coordinates

$$\ddot{R} - R\dot{\varphi}^2 = -\frac{\partial \Phi}{\partial R}, \tag{2}$$

$$R\ddot{\varphi} + 2\dot{R}\dot{\varphi} = -\frac{1}{R}\frac{\partial \Phi}{\partial \varphi}. \tag{3}$$

The $z$ component in any frame is $\ddot{z} = -\partial\Phi/\partial z$, but I consider here motions in the plane of the disc only, hence I neglect the dependence on $z$.

To linearize equations (2) and (3), one needs expansions of $R$, $\varphi$ and $\Phi$ to first order terms. The epicyclic approximation is valid for particles whose trajectories oscillate around circular orbits. For such particles one can write

$$R(t) = R_0 + R_I(t), \tag{4}$$

$$\varphi(t) = \varphi_{00} + \Omega_0 t + \varphi_I(t), \tag{5}$$

$$\Phi(R,\varphi,t) = \Phi_0(R) + \Phi_I(R,\varphi,t), \tag{6}$$

where terms with index $I$ are small to the first order, and second- and higher-order terms were neglected. The parameter $\varphi_{00}$ allows the particle to start from any position angle at time $t=0$, so that $\varphi_0 = \varphi_{00} + \Omega_0 t$. Asymmetry $\Phi_I$ in the potential is small and may be time-dependent. The angular velocity $\Omega_0$ on the circular orbit of radius $R_0$ relates to the potential $\Phi_0$ through the zeroth order of (2): $\Omega_0^2 = (1/R_0)(\partial\Phi_0/\partial R)\,|_{R_0}$. The zeroth order of (3) is identically equal to zero, and the first order corrections to (2) and (3) take respectively forms

$$\ddot{R}_I - 4A\Omega_0 R_I - 2R_0\Omega_0\dot{\varphi}_I = -\frac{\partial \Phi_I}{\partial R}\,|_{R_0,\varphi_0}, \tag{7}$$

$$R_0\ddot{\varphi}_I + 2\Omega_0\dot{R}_I = -\frac{1}{R_0}\frac{\partial \Phi_I}{\partial \varphi}\,|_{R_0,\varphi_0}, \tag{8}$$

where $A$ is the Oort constant defined by $4A\Omega_0 = \Omega_0^2 - \frac{\partial^2\Phi_0}{\partial R^2}|_{R_0}$.

We assume that the bars are point-symmetric with respect to the galaxy centre. Thus to first order the departure of the barred potential from axial symmetry can be described by a term $\cos(2\varphi)$. If multiple bars, indexed by $i$, rotate independently as solid bodies with angular velocities $\Omega_i$, the time-dependent first-order correction $\Phi_I$ to the potential can be written as

$$\Phi_I(R,\varphi,t) = \sum_i \Psi_i(R)\cos[2(\varphi - \Omega_i t)], \tag{9}$$

where the radial dependence $\Psi_i(R)$ has been separated from the angle dependence. No phase in the trigonometric functions above means that we define $t=0$ when all the bars are aligned. Derivatives of (9) enter right-hand sides of (7) and (8), which after introducing $\omega_i = 2(\Omega_0 - \Omega_i)$ take the form

$$\ddot{R}_I - 4A\Omega_0 R_I - 2R_0\Omega_0\dot{\varphi}_I = -\sum_i \frac{\partial \Psi_i}{\partial R}\,|_{R_0}\cos(\omega_i t + 2\varphi_{00}), \tag{10}$$

$$R_0\ddot{\varphi}_I + 2\Omega_0\dot{R}_I = \frac{2}{R_0}\sum_i \Psi_i(R_0)\sin(\omega_i t + 2\varphi_{00}). \tag{11}$$

In order to solve the set of equations (10,11), one can integrate (11) and get an expression for $R_0\dot{\varphi}_I$, which furthermore can be substituted to (10). This substitution eliminates $\varphi_I$, and one gets a single second order equation for $R_I$, which can be written schematically as

$$\ddot{R}_I + \kappa_0^2 R_I = \sum_i A_i \cos(\omega_i t + 2\varphi_{00}) + C_\varphi, \tag{12}$$

where $A_i = -\frac{4\Omega_0 \Psi_i}{\omega_i R_0} - \frac{\partial \Psi_i}{\partial R}_{|R_0}$, $\kappa_0^2 = 4\Omega_0(\Omega_0 - A)$, and $C_\varphi/2\Omega_0$ is the integration constant that appears after integrating (11). This is the equation of a harmonic oscillator with multiple forcing terms, whose solution is well known. It can be written as

$$R_I(t) = C_1 \cos(\kappa_0 t + \delta) + \sum_i M_i \cos(\omega_i t + 2\varphi_{00}) + C_\varphi/\kappa_0^2. \tag{13}$$

The first term of this solution corresponds to a free oscillation at the local epicyclic frequency $\kappa_0$, and $C_1$ is unconstrained. The terms under the sum describe oscillations resulting from the forcing terms in (9), and $M_i$ are functions of $A_i$. Hereafter I focus on solutions without free oscillations, thus I assume that $C_1 = 0$. These solutions will lead to closed periodic orbits and to loops. The formula for $\varphi_I(t)$ can be obtained by substituting (1) into the time-integrated (11). As a result, one gets

$$\dot{\varphi}_I = \sum_i N_i \cos(\omega_i t + 2\varphi_{00}) - \frac{2AC_\varphi}{\kappa_0^2 R_0}, \tag{14}$$

where again $N_i$ are determined by the coefficients of the equations above. Note that to the first order $\Omega_0[R_0 + C_\varphi/\kappa_0^2] = \Omega_0[R_0] - 2AC_\varphi/\kappa_0^2 R_0$, thus the integration constants entering (1) and (14) correspond to a change in the guiding radius $R_0$, and to the appropriate change in the angular velocity $\Omega_0$. They all can be incorporated into $R_0$, and in effect the unique solutions for $R_I$ and $\varphi_I$ are

$$R_I(t) = \sum_i M_i \cos(\omega_i t + 2\varphi_{00}), \tag{15}$$

$$\varphi_I(t) = \sum_i N_i' \cos(\omega_i t + 2\varphi_{00}) + const, \tag{16}$$

where free oscillations have been neglected. The integration constant in (16) is an unconstrained parameter of the order of $\varphi_I$.

### 2.1 Closed Periodic Orbits in a Single Bar

In a potential with a single bar there is only one term in the sums (15) and (16), hereafter indexed with $B$. Consider the change in values of $R_I$ and $\varphi_I$ for a given particle after half of its period in the frame corotating with the bar. This interval is taken because the bar is bisymmetric, so its forcing is periodic with the period $\pi$ in angle. After replacing $t$ by $t + \pi/(\Omega_0 - \Omega_B)$ one gets

$$\begin{aligned} R_I &= M_B \cos[\omega_B(t + \frac{\pi}{\Omega_0 - \Omega_B}) + 2\varphi_{00}] \\ &= M_B \cos(\omega_B t + 2\pi + 2\varphi_{00}). \end{aligned}$$

Thus the solution for $R_I$ after time $\pi/(\Omega_0 - \Omega_B)$ returns its starting value, and the same holds true for $\varphi_I$. After twice that time, i.e. in a full period of this particle in the bar frame, the epicycle centre returns to its starting point and the orbit closes. Thus (15) and (16) describe closed periodic orbits in the linearized problem of a particle motion in a single bar.

### 2.2 Loops in Double Bars

When two independently rotating bars coexist in a galaxy (hereafter indexed by $B$ and $S$), there is no reference frame in which the potential is constant. Thus when a term from one bar in (15) and (16) returns to its starting value, the term from the other bar does not (unless the frequencies of the bars are commensurate). Therefore the particle's trajectory does not close in any reference frame. However, consider the change in value of $R_I$ and $\varphi_I$ after time $\pi/(\Omega_S - \Omega_B)$, which is the relative period of the bars. One gets

$$\begin{aligned} R_I &= M_B \cos[\omega_B(t + \frac{\pi}{\Omega_S - \Omega_B}) + 2\varphi_{00}] + M_S \cos[\omega_S(t + \frac{\pi}{\Omega_S - \Omega_B}) + 2\varphi_{00}] \\ &= M_B \cos(\omega_B t + 2\pi \frac{\Omega_0 - \Omega_B}{\Omega_S - \Omega_B} + 2\varphi_{00}) + M_S \cos(\omega_S t + 2\pi \frac{\Omega_0 - \Omega_S}{\Omega_S - \Omega_B} + 2\varphi_{00}) \\ &= M_B \cos(\omega_B t + 2\pi + 2\varphi_{01}) + M_S \cos(\omega_S t + 2\varphi_{01}), \end{aligned}$$

where $\varphi_{01} = \varphi_{00} + \pi \frac{\Omega_0 - \Omega_S}{\Omega_S - \Omega_B}$. The same result can be obtained for $\varphi_I$. This means that the time transformation $t \to t + \pi/(\Omega_S - \Omega_B)$ is equivalent to the change in the starting position angle of a particle from $\varphi_{00}$ to $\varphi_{01}$. Consider motion of a set of particles that have the same guiding radius $R_0$, but start at various position angles $\varphi_{00}$. This is a one-parameter set, therefore in the disc plane it is represented by a curve, and because of continuity of (15) and (16) this curve is closed. After time $\pi/(\Omega_S - \Omega_B)$, a particle starting at angle $\varphi_{00}$ will take the place of the particle which started at $\varphi_{01}$, a particle starting at $\varphi_{01}$ will take the place of another particle from this curve and so on. The whole curve will regain its shape and position every $\pi/(\Omega_S - \Omega_B)$ time interval, although positions of particles on the curve will shift. This curve is the epicyclic approximation to the *loop*: a curve made of particles moving in a given potential, such that the curve returns to its original shape and position periodically. In the case of two bars, the period is the relative period of the bars, and the loop is made out of particles having the same guiding radius $R_0$. Particles on the loop respond to the forcing from the two bars, but they lack any free oscillation. An example of a set of loops in a doubly barred galaxy in the epicyclic approximation can be seen in [7]. Since they occupy a significant part of the disc, one should anticipate large zones of ordered motions also in the general, non-linear solution for double bars.

## 3 Full Nonlinear Solution for Loops in Nested Bars

Tools and concepts useful in the search for ordered motions in double bars are best introduced through the inspection of particle trajectories in such systems. For this inspection I chose the potential of Model 1 defined in [8], where the small bar is 60% in size of the big bar, and pattern speeds of the bars are not commensurate. Consider a particle moving in this potential inside the corotation of the small bar. Simple experiments with various initial velocities show that if the initial velocity is small enough, the particle usually remains bound. A typical trajectory is shown in the left panels of Fig.1 – since it depends on the reference frame, it is written twice, for reference frame of each bar. Further experimenting with initial velocities shows that particle trajectories are often even tidier: they look like those in the right panels of Fig.1, as if the trajectories were trapped around some regular orbit.

Fine adjustments of the initial velocity lead to a highly harmonious trajectory (Fig.2), which looks like a loop orbit in a potential of a single bar (see e.g. Fig.3.7a in [19]). This is only a formal similarity, but understanding it will let us find out what kind of orbit we see in Fig.2. The loop orbit in a single bar forms when a particle oscillates around a closed periodic orbit. Therefore two frequencies are involved: the frequency of the free oscillation, and the forcing frequency of the bar. On the other hand, the Fourier transform of the trajectory from Fig.2 shows two sharp peaks at frequencies equal to the forcing frequencies of the two bars (Fig.3). Thus the trajectory from Fig.2 also has two frequencies: this time these are the forcing frequencies from the two bars, while the free oscillation is absent. This is how the solution in the linear approximation (Sect. 2.2) was constructed. We conclude that in both the linear (epicyclic approximation) case

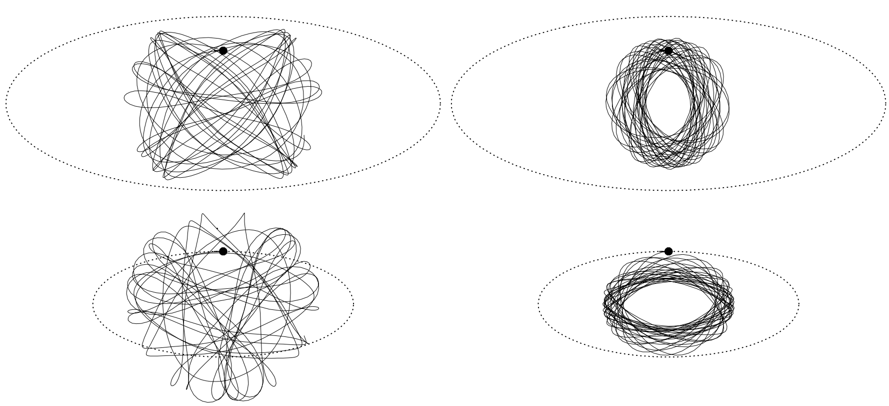


**Fig. 1.** Two example trajectories (one in the two left panels, one in the two right ones) of a particle that moves in the potential of two independently rotating bars. The particle is followed for 10 relative periods of the bars, and its trajectory is displayed in the frame corotating with the big bar (top panels), and the small bar (bottom panels). Each bar is outlined in its own reference frame by the dotted line. Large dot marks the starting point of the particle.

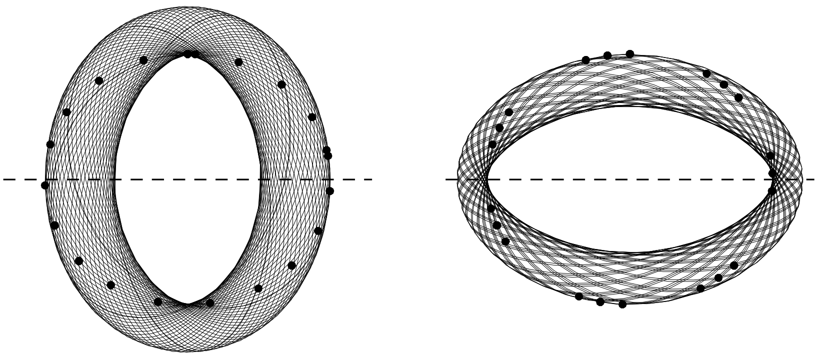


**Fig. 2.** A doubly periodic orbit in the doubly barred potential, followed for 20 relative periods of the bars, and written in the frame corotating with the big bar (left), and the small bar (right). The long axis of each bar is marked by the dashed line. Dots mark positions of the particle at every alignment of the bars.

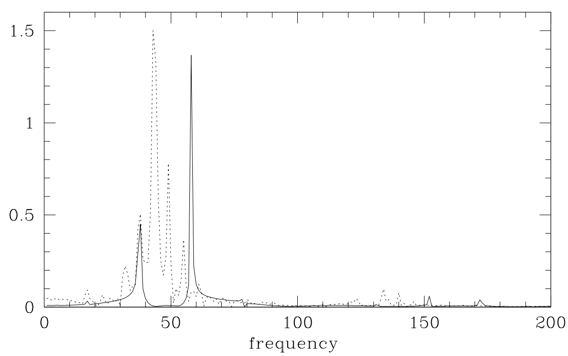



**Fig. 3.** Fourier transforms of the trajectories from right panels of Fig.1 (dotted line) and from Fig.2 (solid line). The peaks in the solid line are related to the forcing frequencies of the bars, and the peaks in the dotted line are not.

and the general case we are dealing with doubly periodic orbits in an oscillating potential of a double bar, with frequencies equal to the forcing frequencies of the bars. In the epicyclic approximation, these orbits have a nice feature that particles following them populate loops: closed curves that return to their original shape and position at every alignment of the bars. One may therefore expect that also in the general case these particles gather on loops.

If in the general case particles on doubly periodic orbits form a loop, one can construct it by writing positions of a particle on such an orbit every time the bars align. These positions are the initial conditions for particles forming the loop, because after every alignment, the $n^{th}$ particle generated in this way takes the position of particle $n+1$. The first 20 positions of a particle on a doubly periodic orbit are overplotted in Fig.2. They indeed seem to be arranged on a closed ellipse-like curve; the shape of this curve varies in time, but it returns to where it started at every alignment of the bars (Fig.4). This construction shows

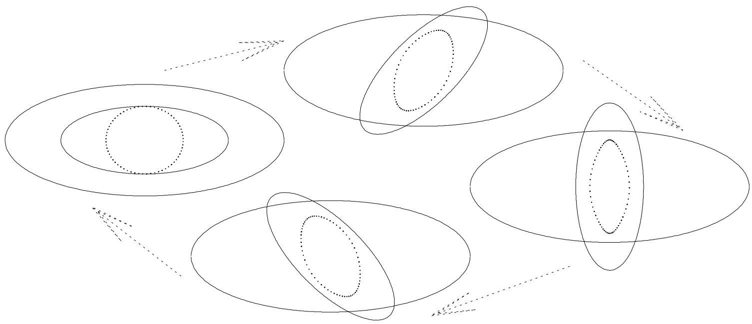


**Fig. 4.** Evolution of the loop from Fig.2 during one relative period of the bars. The bars, outlined with solid lines, rotate counterclockwise. The loop is made out of points that represent separate particles on doubly periodic orbits.

that in the general case particles on doubly periodic orbits also form loops. Note that positions of particles on other orbits, which involve free oscillations, when written at every alignment of the bars, densely populate some two-dimensional section of the plane, and do not gather on any curve. It is extremely useful for the investigation of the orbital structure in double bars that the appearance of the loop is frame-independent. Loops provide an efficient way to classify doubly periodic orbits, which has been hampered so far by the dependence of the last ones on the reference frame.

It turns out that doubly periodic orbits play crucial role in providing orbital support for the pulsating potential of double bars. No closed periodic orbits have been proposed as candidates for the backbone of such a potential. If in a given potential of two bars there are loops that follow the inner bar, and other loops that follow the outer bar, then one may expect that such a potential is dynamically possible. An example of such a potential has been constructed in [8]. The loop from Fig.4 does not follow either bar in its motion, and therefore it is unlikely that it supports the assumed potential. It can be shown that in that potential, there are no loops which could support the two bars. Thus that potential is not self-consistent. This example shows how efficient is the loop approach in rejecting hypothetical doubly barred systems that have no orbital support.

Doubly periodic orbits in double bars are surrounded by regular orbits in the same way as are the closed periodic orbits in a single bar. In both cases, the trapped regular orbits oscillate around the parent orbit. The trajectory from the right panels of Fig.1 is an example of a regular orbit that is trapped around the doubly periodic orbit from Fig.2. How much of the phase space in double bars is occupied by orbits trapped around doubly periodic orbits? It can be examined by launching a particle from e.g. the minor axis of the bar, in the direction perpendicular to this axis, when the bars are aligned. If the particle is trapped, its positions at every alignment of the two bars lie within a ring containing the loop. The width of this ring depends on the particle's position along the minor axis, and on its velocity. It is displayed in Fig.5 for the potential

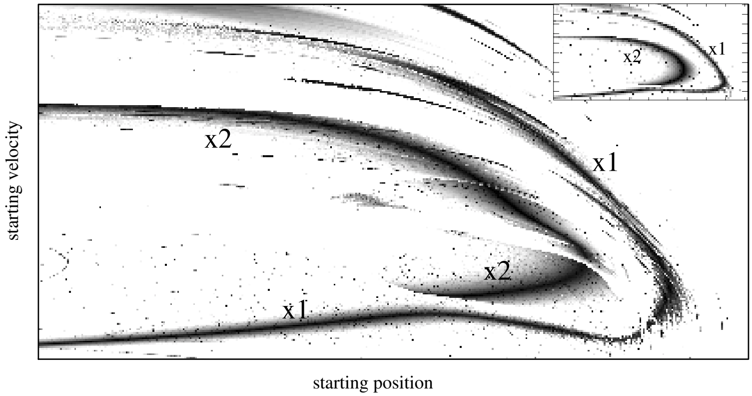



**Fig. 5.** The width of the ring formed by particles trapped around loops in Model 2 from [8] as a function of the particle's position along the minor axis of the aligned bars, and of its velocity (perpendicular to this axis). Darker color means smaller width. In the insert, the same is shown for rings around closed periodic orbits in a single bar (same model, but inner bar axisymmetric). Regions related to the $x_1$ and $x_2$ orbits, and to the loops originating from them, are marked.

of Model 2 defined in [8]. Two stripes of low width appear on the diagram, which correspond to the $x_1$ and $x_2$ orbits in a single bar [3] (displayed in the insert). Thus in double bars there are doubly periodic orbits that correspond to closed periodic orbits in single bars. There are possible regions of chaos in double bars (white stripes in Fig.5), but overall loops in double bars and periodic orbits in single bars trap similar volumes of phase-space around them.

## 4 Conclusions

In a potential of two independently rotating bars, a large fraction of phase space can be occupied by trajectories trapped around parent regular orbits. These orbits are doubly periodic, with the two periods corresponding to the forcing frequencies of the two bars, but they do not close in any reference frame. Like particle trajectories oscillating around closed periodic orbits in a single bar, particle trajectories in double bars oscillate around the doubly periodic parent orbits. The structure of the parent regular orbits can be mapped using the loop approach, which allows us to single out dynamically possible double bars.

### Acknowledgments

The concept of the loop as the organized form of motion in double bars benefits from the insight of Linda Sparke. I thank Lia Athanassoula for our collaboration that lead to this paper, and Peter Erwin for comments on the manuscript.

## References


19. Binney, J. & Tremaine, S. 1987, Galactic Dynamics (Princeton: Princeton Univ. Press)
2. Broucke, R. A. 1969, Periodic Orbits in the Elliptic Restricted Three-Body Problem, Jet Propulsion Laboratory Report 32-1360
3. Contopoulos, G. & Papayannopoulos, Th. 1980, A&A, 92, 33
4. Erwin, P. & Sparke, L. S. 2002, AJ, 124, 65
5. Friedli, D., Wozniak, H., Rieke, M., Martinet, L., & Bratschi, P. 1996, A&AS, 118, 461
6. Louis, P. D. & Gerhard, O. E. 1988, MNRAS, 233, 337
7. Maciejewski, W. & Sparke, L. S. 1997, ApJL, 484, L117
8. Maciejewski, W. & Sparke, L. S. 2000, MNRAS, 313, 745
9. Sridhar, S. 1989, MNRAS, 238, 1159
10. Sygnet, J. F., Tagger, M., Athanassoula, E., & Pellat, R. 1988, MNRAS, 232, 733

# Jeans Solutions for Triaxial Galaxies

Glenn van de Ven[1], Chris Hunter[2], Ellen Verolme[1], and Tim de Zeeuw[1]

[1] Sterrewacht Leiden, Postbus 9513, 2300 RA Leiden, The Netherlands
[2] Department of Mathematics, Florida State University, Tallahassee, FL 32306-4510, USA

**Abstract.** The Jeans equations relate the second-order velocity moments to the density and potential of a stellar system. For general three-dimensional stellar systems, there are three equations, but these are not very helpful, as they contain six independent moments. By assuming that the potential is triaxial and of separable Stäckel form, the mixed moments vanish in confocal ellipsoidal coordinates. The three Jeans equations and three remaining non-vanishing moments form a closed system of three highly-symmetric coupled first-order partial differential equations in three variables. They were first derived by Lynden–Bell in 1960, but have resisted solution by standard methods. Here we present the general solution by superposition of singular solutions.

## 1 Introduction

Much has been learned about the mass distribution and internal dynamics of galaxies by modeling their observed kinematics with solutions of the Jeans equations (e.g. [4]). The Jeans equations connect the second-order velocity moments (or the velocity dispersions, if the mean streaming motion is known) directly to the density and the gravitational potential of the galaxy, without the need to know the phase-space distribution function $f$. In nearly all cases there are fewer Jeans equations than velocity moments, so that additional assumptions have to be made about the degree of anisotropy. Furthermore, the resulting second moments may not correspond to a physical distribution function $f \geq 0$. These significant drawbacks have not prevented wide application of the Jeans approach to the kinematics of spherical and axisymmetric galaxies. Many (components of) galaxies have triaxial shapes ([2], [3]), including early-type bulges, bars, and giant elliptical galaxies. In this geometry, there are three Jeans equations, but little use has been made of them, as they contain six independent second moments, three of which have to be chosen ad-hoc (see e.g. [11]).

An exception is provided by the special set of triaxial mass models that have a gravitational potential of Stäckel form. In these systems, the Hamilton–Jacobi equation separates in confocal ellipsoidal coordinates ([19]), so that all orbits have three exact integrals of motion, which are quadratic in the velocities. The three mixed second-order velocity moments vanish, so that the three Jeans equations for the three remaining second moments form a closed system. Lynden–Bell ([13]) was the first to derive the explicit form of these Jeans equations. He showed that they constitute a highly symmetric set of three first-order partial differential equations for three unknowns, each of which is a function of the ellipsoidal coordinates, but he did not derive solutions.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 101–108, 2003.

When it was realized that the orbital structure in the triaxial Stäckel models is very similar to that in numerical models for triaxial galaxies with cores ([6], [16]), interest in the second moments increased, and the Jeans equations were solved for a number of special cases. These include the axisymmetric limits and elliptic discs ([8], [10]), triaxial galaxies with only thin tube orbits ([12]), and the scale-free limit ([11]). In all these cases the equations simplify to a two-dimensional problem, which can be solved with standard techniques after transforming two first-order equations into a single second-order equation in one dependent variable. However, these techniques do not carry over to a single third-order equation in one dependent variable, which is the best that one could expect to have in the general case. As a result, the latter has remained unsolved.

We have solved the two-dimensional case with an alternative solution method, which does not use the standard approach, but instead uses superposition of singular solutions. This approach can be extended to three dimensions, and provides the general solution for the triaxial case in closed form. We present the detailed solution method elsewhere ([23]), and here we summarise the main results. In ongoing work we will apply our solutions, and will use them together with the mean streaming motions ([20]) to study the properties of the observed velocity and dispersion fields of triaxial galaxies.

## 2 The Jeans Equations for Separable Models

We define confocal ellipsoidal coordinates $(\lambda, \mu, \nu)$ as the three roots for $\tau$ of

$$\frac{x^2}{\tau+\alpha} + \frac{y^2}{\tau+\beta} + \frac{z^2}{\tau+\gamma} = 1 \, , \tag{1}$$

with $(x, y, z)$ the usual Cartesian coordinates, and with constants $\alpha, \beta$ and $\gamma$ such that $-\gamma \le \nu \le -\beta \le \mu \le -\alpha \le \lambda$. Surfaces of constant $\lambda$ are ellipsoids, and surfaces of constant $\mu$ and $\nu$ are hyperboloids of one and two sheets, respectively. The confocal ellipsoidal coordinates are approximately Cartesian near the origin and become conical at large radii, i.e., equivalent to spherical coordinates.

We consider models with a gravitational potential of Stäckel form

$$V_S(\lambda, \mu, \nu) = -\frac{F(\lambda)}{(\lambda-\mu)(\lambda-\nu)} - \frac{F(\mu)}{(\mu-\nu)(\mu-\lambda)} - \frac{F(\nu)}{(\nu-\lambda)(\nu-\mu)} \, , \tag{2}$$

where $F(\tau)$ is an arbitrary smooth function. This potential is the most general form for which the Hamilton–Jacobi equation separates ([15], [18]) All orbits have three exact isolating integrals of motion, which are quadratic in the velocities (e.g. [6]). There are no irregular orbits, so that Jeans' theorem is strictly valid ([14]), and the distribution function $f$ is a function of the three integrals. Therefore, out of the six symmetric second-order velocity moments, defined as

$$\langle v_i v_j \rangle(\boldsymbol{x}) = \frac{1}{\varrho} \iiint v_i v_j f(\boldsymbol{x}, \boldsymbol{v}) \, \mathrm{d}^3 v \, , \quad (i, j = 1, 2, 3), \tag{3}$$

with density $\varrho$, the three mixed moments vanish, and we are left with $\langle v_\lambda^2 \rangle$, $\langle v_\mu^2 \rangle$ and $\langle v_\nu^2 \rangle$, related by three Jeans equations. These were first derived by Lynden–Bell ([13]), and can be written in the following form ([23])

$$\frac{\partial S_{\lambda\lambda}}{\partial \lambda} - \frac{S_{\mu\mu}}{2(\lambda-\mu)} - \frac{S_{\nu\nu}}{2(\lambda-\nu)} = g_1(\lambda,\mu,\nu) \; , \tag{4a}$$

$$\frac{\partial S_{\mu\mu}}{\partial \mu} - \frac{S_{\nu\nu}}{2(\mu-\nu)} - \frac{S_{\lambda\lambda}}{2(\mu-\lambda)} = g_2(\lambda,\mu,\nu) \; , \tag{4b}$$

$$\frac{\partial S_{\nu\nu}}{\partial \nu} - \frac{S_{\lambda\lambda}}{2(\nu-\lambda)} - \frac{S_{\mu\mu}}{2(\nu-\mu)} = g_3(\lambda,\mu,\nu) \; , \tag{4c}$$

where we have defined the diagonal components of the stress tensor

$$S_{\tau\tau}(\lambda,\mu,\nu) = \sqrt{(\lambda-\mu)(\lambda-\nu)(\mu-\nu)}\, \varrho \langle v_\tau^2 \rangle \; , \qquad \tau = \lambda, \mu, \nu, \tag{5}$$

and the functions $g_1$, $g_2$ and $g_3$ depend on the density and potential (2) as

$$g_1(\lambda,\mu,\nu) = -\sqrt{(\lambda-\mu)(\lambda-\nu)(\mu-\nu)}\, \varrho \frac{\partial V_S}{\partial \lambda} \; , \tag{6}$$

where $g_2$ and $g_3$ follow from $g_1$ by cyclic permutation $\lambda \to \mu \to \nu \to \lambda$. Similarly, the three Jeans equations follow from each other by cyclic permutation. The stress components have to satisfy the following continuity conditions

$$S_{\lambda\lambda}(-\alpha,-\alpha,\nu) = S_{\mu\mu}(-\alpha,-\alpha,\nu) \; , \quad S_{\mu\mu}(\lambda,-\beta,-\beta) = S_{\nu\nu}(\lambda,-\beta,-\beta) \; , \tag{7}$$

at the focal ellipse ($\lambda = \mu = -\alpha$) and focal hyperbola ($\mu = \nu = -\beta$), respectively.

We prefer the form (5) for the stresses instead of the more common definition without the square root, since it results in more convenient and compact expressions. In self-consistent models, the density $\varrho$ equals $\varrho_S$, with $\varrho_S$ related to $V_S$ by Poisson's equation. The Jeans equations, however, do not require self-consistency, so that we make no assumptions on the form of $\varrho$ other than that it is triaxial, i.e., a function of $(\lambda, \mu, \nu)$, and that it tends to zero at infinity.

## 3 The Two-Dimensional Case

When two or all three of the constants $\alpha$, $\beta$ or $\gamma$ in (1) are equal, the triaxial Stäckel models reduce to limiting cases with more symmetry and thus with fewer degrees of freedom. Solving the Jeans equations for oblate, prolate, elliptic disc and scale-free models reduces to the same two-dimensional problem ([10], [11], [23]), of which the simplest form is the pair of Jeans equations for Stäckel discs

$$\frac{\partial S_{\lambda\lambda}}{\partial \lambda} - \frac{S_{\mu\mu}}{2(\lambda-\mu)} = g_1(\lambda,\mu) \; , \tag{8a}$$

$$\frac{\partial S_{\mu\mu}}{\partial \mu} - \frac{S_{\lambda\lambda}}{2(\mu-\lambda)} = g_2(\lambda,\mu) \; , \tag{8b}$$

with at the foci ($\lambda = \mu = -\alpha$) the continuity condition

$$S_{\lambda\lambda}(-\alpha,-\alpha) = S_{\mu\mu}(-\alpha,-\alpha) \, . \tag{9}$$

In this case the stress components and the functions $g_1$ and $g_2$ are

$$S_{\tau\tau}(\lambda,\mu) = \sqrt{(\lambda-\mu)}\,\varrho\langle v_\tau^2\rangle \quad (\tau=\lambda,\mu), \quad g_1(\lambda,\mu) = -\sqrt{(\lambda-\mu)}\,\varrho\,\frac{\partial V_S}{\partial\lambda} \, , \tag{10}$$

where $g_2$ follows from $g_1$ by interchanging $\lambda \leftrightarrow \mu$, and $\varrho$ denotes a surface density.

The two Jeans equations (8) can be recast into a single second-order partial differential equation in either $S_{\lambda\lambda}$ or $S_{\mu\mu}$, which can be solved by employing standard techniques like Riemann's method ([5], [23]). However, these standard techniques do not carry over to the triaxial case, and we therefore introduce an alternative method, based on the superposition of singular solutions.

We consider a simpler form of (8) by substituting for $g_1$ and $g_2$, respectively $\tilde{g}_1 = 0$ and $\tilde{g}_2 = \delta(\lambda_0-\lambda)\delta(\mu_0-\mu)$. We refer to solutions of these simplified Jeans equations as *singular solutions*. Singular solutions can be interpreted as contributions to the stresses at a fixed field point $(\lambda,\mu)$ due to a source point in $(\lambda_0,\mu_0)$ (Fig. 1). The full stress at the field point can be obtained by adding all source point contributions, each with a weight that depends on the local density and potential. Once we know the singular solutions, we can use the superposition principle to construct the the solution of the full Jeans equations (8).

Since the derivative of a step-function $\mathcal{H}$ is equal to a delta-function, it follows that the singular solutions must have the form

$$\begin{aligned} S_{\lambda\lambda} &= A(\lambda,\mu)\mathcal{H}(\lambda_0-\lambda)\mathcal{H}(\mu_0-\mu) \, , \\ S_{\mu\mu} &= B(\lambda,\mu)\mathcal{H}(\lambda_0-\lambda)\mathcal{H}(\mu_0-\mu) - \delta(\lambda_0-\lambda)\mathcal{H}(\mu_0-\mu) \, . \end{aligned} \tag{11}$$

where the functions $A$ and $B$ must solve the homogeneous Jeans equations, i.e., (8) with zero right-hand side, and satisfy the following boundary conditions

$$A(\lambda_0,\mu) = \frac{1}{2(\lambda_0-\mu)} \, , \quad B(\lambda,\mu_0) = 0 \, . \tag{12}$$

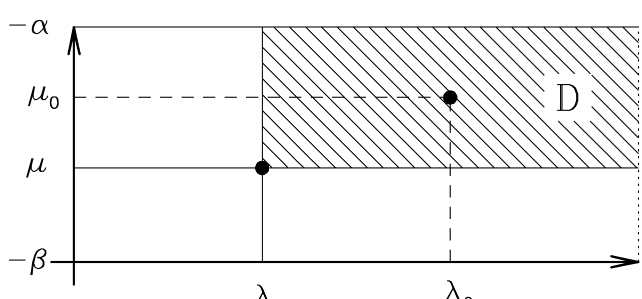



**Fig. 1.** *The $(\lambda_0,\mu_0)$-plane. The total stress at a field point $(\lambda,\mu)$ consists of the weighted contributions from source points at $(\lambda_0,\mu_0)$ in the domain $D$.*

We solve this two-dimensional homogeneous boundary problem by superposition of particular solutions. We first derive a particular solution of the homogeneous Jeans equations with a free parameter $z$, which we assume to be complex. We then construct a linear combination of these particular solutions by integrating over $z$. We choose the integration contours in the complex $z$-plane, such that the boundary conditions (12) are satisfied simultaneously. The resulting homogeneous solutions are complex contour integrals, which can be evaluated in terms of the complete elliptic integral of the second kind, $E(w) \equiv \int_0^{\frac{\pi}{2}} \mathrm{d}\theta \sqrt{1 - w \sin^2 \theta}$, and its derivative $E'(w)$, as

$$A = \frac{E(w)}{\pi(\lambda_0 - \mu)}, \quad B = -\frac{2wE'(w)}{\pi(\lambda_0 - \lambda)}, \quad \text{with} \quad w = \frac{(\lambda_0 - \lambda)(\mu_0 - \mu)}{(\lambda_0 - \mu_0)(\lambda - \mu)}. \tag{13}$$

We obtain a similar system of simplified Jeans equations by interchanging the expressions for $\tilde{g}_1$ and $\tilde{g}_2$. The singular solutions of this simplified system follow from (11) by interchanging $\lambda \leftrightarrow \mu$ and $\lambda_0 \leftrightarrow \mu_0$ at the same time.

To find the solution to the full Jeans equations (8) at $(\lambda, \mu)$, we multiply the latter singular solutions and (11) by $g_1(\lambda_0, \mu_0)$ and $g_2(\lambda_0, \mu_0)$ respectively, and integrate over $D = \{(\lambda_0, \mu_0)\colon \lambda \leq \lambda_0 \leq \infty, \mu \leq \mu_0 \leq -\alpha\}$ (Fig. 1). This gives the first two integrals of the two equations (14a) and (14b) below. The remaining terms are due to the non-vanishing stress at the boundary $\mu = -\alpha$, and are found by multiplying the singular solutions (11), evaluated at $\mu_0 = -\alpha$, by $-S_{\mu\mu}(\lambda_0, -\alpha)$ and integrating over $\lambda_0$ in $D$. The final result for the solution of the Jeans equations (8) for Stäckel discs, after using the evaluations (13), is

$$\begin{aligned} S_{\lambda\lambda}(\lambda, \mu) = & \int_\lambda^\infty \mathrm{d}\lambda_0 \int_\mu^{-\alpha} \mathrm{d}\mu_0 \left[ -g_1(\lambda_0, \mu_0) \frac{2wE'(w)}{\pi(\mu_0 - \mu)} + g_2(\lambda_0, \mu_0) \frac{E(w)}{\pi(\lambda_0 - \mu)} \right] \\ & - \int_\lambda^\infty \mathrm{d}\lambda_0 \, g_1(\lambda_0, \mu) - \int_\lambda^\infty \mathrm{d}\lambda_0 \, S_{\mu\mu}(\lambda_0, -\alpha) \left[ \frac{E(w)}{\pi(\lambda_0 - \mu)} \right]_{\mu_0 = -\alpha}, \end{aligned} \tag{14a}$$

$$\begin{aligned} S_{\mu\mu}(\lambda, \mu) = & \int_\lambda^\infty \mathrm{d}\lambda_0 \int_\mu^{-\alpha} \mathrm{d}\mu_0 \left[ -g_1(\lambda_0, \mu_0) \frac{E(w)}{\pi(\lambda - \mu_0)} - g_2(\lambda_0, \mu_0) \frac{2wE'(w)}{\pi(\lambda_0 - \lambda)} \right] \\ & - \int_\mu^{-\alpha} \mathrm{d}\mu_0 \, g_2(\lambda, \mu_0) + S_{\mu\mu}(\lambda, -\alpha) - \int_\lambda^\infty \mathrm{d}\lambda_0 \, S_{\mu\mu}(\lambda_0, -\alpha) \left[ -\frac{2wE'(w)}{\pi(\lambda_0 - \lambda)} \right]_{\mu_0 = -\alpha}. \end{aligned} \tag{14b}$$

The solution depends on $\varrho$ and $V_S$ through $g_1$ and $g_2$. This means that, for given $\varrho$ and $V_S$, the solution is uniquely determined once we have prescribed $S_{\mu\mu}$ at the boundary $\mu = -\alpha$. At this boundary, $S_{\lambda\lambda}$ is related to $S_{\mu\mu}$ by the first Jeans equation (8a), evaluated at $\mu = -\alpha$, up to an integration constant, which is fixed by the continuity condition (9). We are thus free to specify either of the two stress components at $\mu = -\alpha$.

## 4 The General Case

The singular solution method introduced in the previous section can be extended to three dimensions to solve the Jeans equations (4) for triaxial Stäckel models. Although the calculations are more complex for a triaxial model, the stepwise solution method is similar to that in two dimensions.

We simplify the Jeans equations (4) by setting two of the three functions $g_1$, $g_2$ and $g_3$ to zero and the remaining equal to $\delta(\lambda_0-\lambda)\delta(\mu_0-\mu)\delta(\nu_0-\nu)$. In this way, we obtain three similar simplified systems ($i=1,2,3$), each with three singular solutions $S_i^{\tau\tau}(\lambda,\mu,\nu;\lambda_0,\mu_0,\nu_0)$ ($\tau=\lambda,\mu,\nu$), that describe the stress components at a fixed field point $(\lambda,\mu,\nu)$ due to a source point in $(\lambda_0,\mu_0,\nu_0)$.

The singular solutions have a form that is similar to that in the two-dimensional case (11). They consist of combinations of step-functions and delta-functions multiplied by functions that are the solutions of homogeneous boundary problems. The functions that must solve a two-dimensional homogeneous boundary problem can be found as in Sect. 3, and can be expressed in terms of complete elliptic integrals, cf. (13). The singular solutions in the general case also contain three functions $A$, $B$ and $C$ that must solve the triaxial homogeneous Jeans equations, i.e., (4) with zero right-hand side, and satisfy three boundary conditions. This three-dimensional homogeneous boundary problem can be solved by integrating a *two*-parameter particular solution over both its complex parameters, and choosing the combination of contours such that the three boundary conditions are satisfied simultaneously. The resulting homogeneous solutions $A$, $B$ and $C$ are products of complex contour integrals, and can be evaluated as sums of products of complete *hyper*elliptic integrals.

To find the solution of the full Jeans equations (4) , we multiply each singular solution $S_i^{\tau\tau}$ by $g_i(\lambda_0,\mu_0,\nu_0)$, so that the contribution from the source point naturally depends on the local density and potential. Then, for each coordinate $\tau=\lambda,\mu,\nu$, we add the three weighted singular solutions, and integrate over a finite volume within the physical region $-\gamma\le\nu\le-\beta\le\mu\le-\alpha\le\lambda$. This results in the following general solution of the Jeans equations (4) for triaxial Stäckel models

$$
\begin{aligned}
S_{\tau\tau}(\lambda,\mu,\nu) = {} & \int_{\lambda}^{\lambda_e}\mathrm{d}\lambda_0\int_{\mu}^{\mu_e}\mathrm{d}\mu_0\int_{\nu}^{\nu_e}\mathrm{d}\nu_0\sum_{i=1}^{3} g_i(\lambda_0,\mu_0,\nu_0)\,S_i^{\tau\tau}(\lambda,\mu,\nu;\lambda_0,\mu_0,\nu_0)\\
& -\int_{\mu}^{\mu_e}\mathrm{d}\mu_0\int_{\nu}^{\nu_e}\mathrm{d}\nu_0\,S_{\lambda\lambda}(\lambda_e,\mu_0,\nu_0)\,S_1^{\tau\tau}(\lambda,\mu,\nu;\lambda_e,\mu_0,\nu_0)\\
& -\int_{\nu}^{\nu_e}\mathrm{d}\nu_0\int_{\lambda}^{\lambda_e}\mathrm{d}\lambda_0\,S_{\mu\mu}(\lambda_0,\mu_e,\nu_0)\,S_2^{\tau\tau}(\lambda,\mu,\nu;\lambda_0,\mu_e,\nu_0)\\
& -\int_{\lambda}^{\lambda_e}\mathrm{d}\lambda_0\int_{\mu}^{\mu_e}\mathrm{d}\mu_0\,S_{\nu\nu}(\lambda_0,\mu_0,\nu_e)\,S_3^{\tau\tau}(\lambda,\mu,\nu;\lambda_0,\mu_0,\nu_e)\ ,
\end{aligned}
\tag{15}
$$

with $\tau = \lambda, \mu, \nu$. Whereas the integration limits $\lambda$, $\mu$ and $\nu$ are fixed due to the position of the field point, the limits $\lambda_e$, $\mu_e$ and $\nu_e$ are not, and may be any value in the corresponding physical ranges, i.e., $\lambda_e \in [-\alpha, \infty]$, $\mu_e \in [-\beta, -\alpha]$ and $\nu_e \in [-\gamma, -\beta]$, but $\lambda_e \neq -\alpha$. The latter choice would lead to solutions which generally have the incorrect radial fall-off, and hence are non-physical. If we choose $\lambda_e = \infty$, there is no contribution from the second line in (15) due to vanishing stress at large distance. If we furthermore take $\mu_e = -\alpha$ and $\nu_e = -\beta$, the integration volume becomes the three-dimensional extension of $D$ (Fig. 1).

Whereas the volume integral in (15) already solves the inhomogeneous Jeans equations (4) , the three area integrals are needed to obtain the correct values at the boundary surfaces $\lambda = \lambda_e$, $\mu = \mu_e$ and $\nu = \nu_e$. On each of these surfaces the three stress components are related by two of the three Jeans equations (4) and the continuity conditions (7). Since the (weight) functions $g_i$ are known for given $\varrho$ and $V_S$, this means that the solution (15) yields all three stresses everywhere in the triaxial model, once one of the stress components is prescribed on the three boundary surfaces. If we take $\lambda_e = \infty$ and $\mu_e = \nu_e = -\beta$, the contributing boundary surfaces reduce to the single $(x, z)$-plane, containing the long and the short axis of the galaxy. This compares well with Schwarzschild ([17]), who used the same plane to start his numerically calculated orbits from.

## 5 Discussion and Conclusions

Eddington ([9]) showed that the velocity ellipsoid in a triaxial galaxy with a separable potential of Stäckel form is everywhere aligned with the confocal ellipsoidal coordinate system in which the equations of motion separate. Lynden–Bell ([13]) derived the three Jeans equations which relate the three principal stresses to the potential and the density. Solutions were found for the various two-dimensional limiting cases, but with methods that do not carry over to the general case, which remained unsolved. We have presented an alternative solution method, based on the superposition of singular solutions (see [23] for details). This approach, unlike the standard techniques, can be generalised to solve the three-dimensional system. The resulting solutions contain complete (hyper)elliptic integrals, which can be evaluated in a straightforward way.

The general Jeans solution is not unique, but requires specification of principal stresses at certain boundary surfaces, given a separable triaxial potential and a triaxial density distribution (not necessarily the one that generates the potential). These boundary surfaces can be taken to be the plane containing the long and the short axis of the galaxy, and, more specifically, the part that is crossed by all three families of tube orbits and the box orbits.

The set of all Jeans solutions (15) contains all the stresses that are associated with the physical distribution functions $f \geq 0$, but, as in the case of spherical and axisymmetric models, also contains solutions which are unphysical, e.g., those associated with distribution functions that are negative in some parts of phase space. The many examples of the use of spherical and axisymmetric Jeans models in the literature suggest nevertheless that the Jeans solutions can be of significant use.

While triaxial models with a separable potential do not provide an adequate description of the nuclei of galaxies with cusped luminosity profiles and a massive central black hole ([7]), they do catch much of the orbital structure at larger radii, and in some cases even provide a good approximation of the galaxy potential. The solutions for the mean streaming motions, i.e., the first velocity moments of the distribution function, are helpful in understanding the variety of observed velocity fields in giant elliptical galaxies and constraining their intrinsic shapes (e.g. [1], [21], [22]). We expect that the projected velocity dispersion fields that can be derived from our Jeans solutions will be similarly useful, and, in particular, that they can be used to establish which combinations of viewing directions and intrinsic axis ratios are firmly ruled out by the observations.

It is remarkable that the entire Jeans solution can be written down by means of classical methods. This suggests that similar solutions can be found for the higher dimensional analogues of (4) , most likely involving hyperelliptic integrals of higher order. It is also likely that the higher-order velocity moments for the separable triaxial models can be found by similar analytic means, but the effort required may become prohibitive.

## References


1. Arnold R., de Zeeuw P. T., Hunter C., 1994, MNRAS, 271, 924
2. Binney J., 1976, MNRAS, 177, 19
3. Binney J., 1978, MNRAS, 183, 501
4. Binney J., Tremaine S., 1987, Galactic Dynamics. Princeton, NJ, Princeton University Press
5. Copson E. T., 1975, Partial Differential Equations. Cambridge, Cambridge University Press
6. de Zeeuw P. T., 1985, MNRAS, 216, 273
7. de Zeeuw P. T., Peletier R., Franx M., 1986, MNRAS, 221, 1001
8. Dejonghe H., de Zeeuw P. T., 1988, ApJ, 333, 90
9. Eddington A. S., 1915, MNRAS, 76, 37
10. Evans N. W., Lynden-Bell D., 1989, MNRAS, 236, 801
11. Evans N. W., Carollo C. M., de Zeeuw P. T., 2000, MNRAS, 318, 1131
12. Hunter C., de Zeeuw P. T., 1992, ApJ, 389, 79
13. Lynden-Bell D., 1960, PhD thesis, Cambridge University
14. Lynden-Bell D., 1962, MNRAS, 124, 1
15. Lynden-Bell D., 1962, MNRAS, 124, 95
16. Schwarzschild M., 1979, ApJ, 232, 236
17. Schwarzschild M., 1993, ApJ, 409, 563
18. Stäckel P., 1890, Math. Ann., 35, 91
19. Stäckel P., 1891, Über die Integration der Hamilton-Jacobischen Differential gleichung mittelst Separation der Variabeln. Habilitationsschrift, Halle
20. Statler T. S., 1994, ApJ, 425, 458
21. Statler T. S., 2001, AJ, 121, 244
22. Statler, T. S., Dejonghe, H., Smecker-Hane, 1999, AJ, 117, 126
23. van de Ven G., Hunter C., Verolme E. K., de Zeeuw P. T., 2003, MNRAS, submitted

# Nonlinear Response of the Interstellar Gas Flow to Galactic Spiral Density Waves

Edward Liverts[1,2] and Michael Mond[1]

[1] Ben-Gurion University, of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel
[2] Yaroslav-the-Wise Novgorod State University, Novgorod 173021, Russia

**Abstract.** Supersonic nonlinear gas flow is studied in order to describe galactic spiral density waves. It is shown analytically that ultra harmonic periodic solutions may exist if nonlinear effects are taken into account. The relevance of those solutions to observed data is discussed.

## 1 Introduction

An extensive literature has been devoted to simulations of the large-scale flow of interstellar gas in a stellar spiral density wave in galaxies, e.g. [1–4]. This study has played an important role in the understanding of many processes occurring in galaxies such as the structure of dust lanes observed along the inner edges of spiral arms in many galaxies [5, 6]; the enhanced synchrotron radiation from spiral arms [7, 8]; the radio emission of HI at the wavelength 21 cm [9]; a trigger mechanism for star formation, and the creation the narrow bands of young highly luminous stars [2, 10, 11]. A review of the problem is presented in [12–14]. It has been shown that spiral density waves, propagating in the galactic disk and interpreted as the galactic spiral arms, may induce large nonlinear perturbations in the gas flow. It was suggested that such nonlinear phenomena take place in the gas even though the amplitude of the spiral stellar field is relatively small. This is so since the response to the gravitational potential induced by the stellar density wave is roughly proportional to $a^{-2}$, where $a$ is the velocity dispersion for stars, and the sound speed for the gas [4]. For the gas $a \sim 8$ km/s, while for the stars $a \sim 40$ km/s. Hence, if the amplitude of the perturbation field is small, one can use a linear theory to calculate a disturbance in the stellar disk. In contrast, the perturbation in the gaseous component is much stronger and one has to use a nonlinear theory. The steady flow in the nonlinear regime was considered numerically and it was shown, in particular, that a secondary shock wave is possible (this was borne out by numerical calculations carried out for the range of galactocentric radius $r$ from $10.0kpc$ to $12.5kpc$ in the adopted model) see [11]. The nonlinear effect may well account for the origin of the Sagitta-Carina feature and relatively short spurs or feathers (see [11], and references therein). Moreover, the nonlinear effect may provide an answer to the old puzzle of how a two-arm potential drives multiple arms. In optical images we can see primarily a brightness distribution, which, generally speaking, does not reflect the over-all mass distribution. The spiral arms owe their high luminosity to the fact that the brightest objects in the galaxy are concentrated in them: giant stars

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 109–116, 2003.

of the early OB spectral classes and ionized hydrogen HII regions which have high luminosity in their emission line. Certain galaxies show a distinct spiral structure outlined by stars of later spectral class that belong to the old population of galaxy's spherical and disk subsystems. This effect in galactic morphology was first noticed by Zwicky [15] who detected smooth red arms in the disk of the grand design galaxy M51 and showed that the morphology of the evolved disk population need not follow the Hubble classification assigned from the young population tracers. A similar point was also made by Vorontsov-Vel'yaminov [16] and the same picture was later observed in certain other galaxies. Two images of the giant grand design Hubble type Sc galaxy NGC 309 seen almost face-on, one in blue light and the other in near infrared were published [17].

The aim of this paper is to analytically obtain conditions under which spiral density wave may give rise to a nonlinear response of parametrically excited and forced gas flow. That question is considered in Sect. 3 In addition it is aimed to consider gas-star structures presented in Sect. 4 that are typical for real galaxies in contrast to the results obtained from the purely gaseous response calculations without taking into account the nonlinear effects.

## 2 Basic Equations

Consider the flow of a galactic interstellar gas under the influence the gravitational potential due to the galactic stars. It is assumed that the gas rotates with a given angular velocity $\Omega(r)$ at a distance $r$ from the galactic center, while the stars are assumed to be arranged along two spiral arms that result from a density wave with an angular phase velocity $\Omega_p$. Hence, in a frame that rotates with angular velocity $\Omega_p$, the arms appear stationary. As a result, it is convenient to write the hydrodynamic equations that describe the steady gas flow in a frame that rotates with the angular velocity $\Omega_p$. They are:

$$\nabla \cdot (\rho \mathbf{v}) = 0, \tag{1}$$

$$(\mathbf{v} \cdot \nabla)\mathbf{v} + 2\boldsymbol{\Omega}_p \times \mathbf{v} = -\frac{1}{\rho}\nabla P - \nabla(\Phi - \frac{1}{2}{\Omega_p}^2 r^2). \tag{2}$$

In (1)–(2), $\rho$, $\mathbf{v}$, and $P = a^2\rho$ are the density, velocity, and pressure of the interstellar gas, respectively, and $\Phi$ is the gravitational potential of the stellar subsystem.

It is assumed that the potential $\Phi$ is given by the sum of an axisymmetric unperturbed potential $\Phi_0(r)$ and a non axisymmetric perturbation $\Phi_1(\mathbf{r}, t)$. The non axisymmetric part of gravitational potential is due to the stellar spiral density wave and is of the form [2]

$$\Phi_1(\mathbf{r}, t) = \Re\bar{\phi}_1 e^{i(2(\Omega_p t - \varphi) + \int k(r)dr)} \tag{3}$$

where $k$ is the radial wave number of the two-armed trailing spiral wave and is related to the angle of inclination of a spiral arm to the circumferential direction $i$ by the relation $k = -2/r/\tan i$. Here and below $r$, $\varphi$ are the galactocentric cylindrical coordinates and the axis of the galactic rotation is along the $z$-axis. Obviously, the characteristic length in this problem is $1/k \sim \lambda$. According to observations and following [2] a tightly wound spiral perturbation is considered, in which the angle of inclination $i$ is small, hence $\lambda/r \ll 1$ or $|k|r \gg 1$. It is useful to introduce the spiral coordinates $(\eta, \xi)$. The spiral coordinates are fixed in the rotating frame (with angular velocity $\Omega_p$) such that $\eta$ is constant along the spiral arms while $\xi$ is constant along lines that are orthogonal to the spiral arms. Using the above definition of the spiral coordinates, the perturbed spiral stellar gravitational potential can be written in the form [2]

$$\Phi_1(\eta) = F\frac{\Omega^2 r\cos^2 i}{k}\cos\eta, \tag{4}$$

where $F$ is the amplitude of the perturbation due to the stellar spiral density wave as a fraction of the unperturbed axisymmetric gravitational potential. The gaseous response depends on the strength of the density wave gravitational field [11, 20] and have to be considered as nonlinear perturbation even though the moderate spiral forcing being 10% of the axisymmetric force field [19, 20]

We turn now to deriving the equations that describe the gas flow under the stellar gravitational potential in the spiral coordinate system defined above. In order to do that we assume that the gas flow is given by an axisymmetric basic flow that describes the response of the gas to the axisymmetric unperturbed potential $\Phi_0$ plus a perturbation which results from the spiral perturbation in the stellar gravitational potential $\Phi_1$, as given in (4).

The solutions of (1) and (2) for rapidly rotating thin gaseous disk that describes the basic axisymmetric flow is readily obtained in the spiral coordinate system as

$$v_{||0} = (\Omega-\Omega_p)r\cos i, \qquad v_{\perp 0} = (\Omega-\Omega_p)r\sin i = \frac{2}{k}(\Omega-\Omega_p)\cos i, \tag{5}$$

where $v_{||}, v_{\perp}$ are the gas velocity components parallel and perpendicular to the spiral arms, respectively. It is easy to see that $|v_{\perp 0}/v_{||0}| \sim 1/|k|r \ll 1$.

In order to describe the perturbed variables the following angular velocity scale is introduced:

$$\chi = 2\Omega[1+\frac{r}{2\Omega}\frac{d\Omega}{dr}\cos^2 i]^{1/2}, \tag{6}$$

with the aid of which the following normalization is introduced for the velocity components:

$$u = \frac{v_{\perp 1}}{\bar{v}_{\perp}}, \quad \bar{v}_{\perp} = \chi\frac{\cos i}{k}, \quad v = \frac{v_{||1}}{\bar{v}_{||}}, \quad \bar{v}_{||} = \frac{\chi^2}{\Omega}\frac{\cos i}{k} \tag{7}$$

Example for values for the various parameters can be estimated from the galactic equilibrium parameters at the solar position as given in [11]:

$$\Omega = 24.7\frac{km}{sec\ kpc}, \quad \Omega_p = 13.5\frac{km}{sec\ kpc}, \quad \chi = 31\frac{km}{sec\ kpc}. \tag{8}$$

Inserting (4-7) into (1) and (2), eliminating the perturbed density $\rho_1$, and assuming that the derivatives across the spiral arms are much bigger than the derivatives along the arms [18] result in the following two equations for the perturbed velocity components:

$$\frac{(-\nu+u)^2-c^2}{-\nu+u}\,\frac{du}{d\eta}=v-f\sin\ \eta,\quad u+(-\nu+u)\frac{dv}{d\eta}=0, \tag{9}$$

where

$$f=F(\Omega^2/\chi^2)kr,\quad c^2=\frac{a^2k^2}{\chi^2\cos^2 i},\quad \nu=-\frac{v_{\perp 0}}{\bar{v}_{\perp}}, \tag{10}$$

are the dimensionless amplitude of the stellar density spiral wave, the squared dimensionless sound speed, and the dimensionless basic perpendicular velocity, respectively. The values for those parameters as estimated from (8) are $0.1-0.2$, 0.195, and 0.72, respectively.

Finally, by eliminating $v$ and defining a new variable $y=u/(f\nu)$, a single second order ordinary differential equation is obtained:

$$\begin{aligned}y''&+\omega_0^2(1+2f\cos\eta-f^2y\cos\eta)y=\omega_0^2\cos\eta+f((2\nu^2\omega_0^2+1)yy''+\\&+(2\nu^2\omega_0^2-1)y'^2+\omega_0^2y^2)+f^2\nu^2\omega_0^2(-3y^2y''-2yy'^2)+\\&+f^3\nu^2\omega_0^2(y^3y''+y^2y'^2).\end{aligned} \tag{11}$$

where $\omega_0^2=(\nu^2-c^2)^{-1}$ is the natural frequency of linear oscillations near steady state as a result of a small imbalance between the Coriolis force and the gaseous pressure.

## 3 Ultra Harmonic Resonances

The solution of(11) for the cases in which $\omega_0 \not\approx 2,3,\ldots,n$ was presented in [20]. However, the response of the $n-th$ harmonic (ultra harmonic resonances response) can be expected to be sufficiently large if $\omega_0\approx 2,3,\ldots,n$ as a result of the combined ultra harmonic and the parametric resonances. The existence of this ultra harmonic resonances has been recently demonstrated by the numerical calculations [11]. Equation (11) will be analyzed analytically in this section by the method of multiple scales which is often used in the analysis of weakly nonlinear dynamic systems [21].

### 3.1 The Resonant Case of $\omega_0 \approx 2$

In this the combined ultra harmonic and the parametric resonances response when $\omega_0\approx 2$ will be studied. The solution of (11) to first order in the small parameter $f$ is straight forward and is given by

$$y_0=a_0\cos(\omega\eta)+2\Lambda\cos\eta,\quad \Lambda=\frac{1}{2}\frac{\omega_0^2}{\omega_0^2-1} \tag{12}$$

Substituting (12) into (11), equating coefficient of like powers of $f$, one obtains an equation for determining the next-order terms of the expansion of the solution $y(\eta)$. However, the fact that $\omega_0$ is close to 2 gives rise to secular terms and small-divisors in the higher order equations. Consequently, in order to proceed further, a detuning parameter $\sigma$ is introduced

$$\omega_0 = 2 + \sigma, \tag{13}$$

as well as higher order deviations of the frequency from $\omega_0$

$$\omega = \omega_0 + \omega^{(1)} + \omega^{(2)} + \dots. \tag{14}$$

Using the above definitions, the equation for $y_1$ may be rewritten as:

$$\begin{aligned} y_1'' + \omega_0^2 y_1 &= f(-8\nu^2\omega_0^2\Lambda^2 + 2\omega_0^2\Lambda(\Lambda - 1))\cos 2\eta \\ +2\omega_0 a_0' \sin(\omega_0 + \omega^{(1)})\eta &+ 2\omega_0 a_0 (\omega^{(1)}\eta)' \cos(\omega_0 + \omega^{(1)})\eta + NST \end{aligned} \tag{15}$$

where $NST$ stands for non secular terms. Eliminating the secular terms from last equation yields

$$\begin{aligned} f(-8\nu^2\omega_0^2\Lambda^2 + 2\omega_0^2\Lambda(\Lambda - 1))\sin(\sigma + \omega^{(1)})\eta + 2\omega_0 a_0' &= 0 \\ f(-8\nu^2\omega_0^2\Lambda^2 + 2\omega_0^2\Lambda(\Lambda - 1))\cos(\sigma + \omega^{(1)})\eta + 2\omega_0 a_0 (\omega^{(1)}\eta)' &= 0 \end{aligned} \tag{16}$$

As $a_0$ as well as $\omega^{(1)}$ are constants, it follows from (16) that

$$\omega^{(1)} = -\sigma. \tag{17}$$

It can be seen that unlike the non resonant response for which the amplitude of 2-nd harmonic is proportional to $f$ which is a small parameter, here this amplitude is of order $f/\sigma$ which is of order one. Thus, small perturbations in the stellar gravitational potential give rise to finite response in the gas flow.

### 3.2 The Resonant Case of $\omega_0 \approx 3$

In the case of the combined ultra harmonic and parametric resonances response when $\omega_0 \approx 3$ the small-divisor terms that result from the non linear terms occur at $O(f^2)$ and the amplitudes of the responses have been ordered so that affects of the resonances first occur at $O(f^2)$. The solution of (11) to order $O(f)$ is given by

$$y_1 = a_{10} + 2\Lambda\cos\eta + a_{12}\cos 2\eta + a_{13}\cos\omega\eta, \tag{18}$$

where

$$a_{10} = -2f\Lambda^2, \quad a_{12} = -4f\frac{\Lambda^2}{\omega_0^2 - 4}(2\nu^2\omega_0^2 + \frac{\omega_0^2}{2} - 1) \tag{19}$$

Substituting the solution (18) into (11) yields

$$y_2'' + \omega_0^2 y_2 = A\cos 3\eta + B\cos\omega\eta + 2\omega_0 a_{13}'\sin\omega\eta + \\ +2\omega_0 a_{13}(\omega^{(1)}\eta)'\cos\omega\eta + NST \tag{20}$$

in which

$$A = f\omega_0^2 a_{12}(2\Lambda - 2\nu^2 - 2 - 8\nu^2\Lambda) + f^2\Lambda^2(1 + 10\nu^2\omega_0^2\Lambda)$$
$$B = f(-(2\nu^2\omega_0^2 + 1)\omega^2 a_{10}a_{13} + f\nu^2\omega_0^2(8\Lambda^2 a_{13} + \frac{7}{4}a_{13}{}^3\omega^2))$$

To proceed further, once again a detuning parameter $\sigma$ and the deviation of the frequency from $\omega_0$ due to non-linearity are introduced according to

$$\omega_0 = 3 + \sigma$$
$$\omega = \omega_0 + \omega^{(1)} + \omega^{(2)} + \dots \tag{21}$$

Thus, eliminating the secular terms from (20) yields

$$A\sin(\sigma + \omega^{(1)}) + 2\omega_0 a_{13}' = 0$$
$$A\cos(\sigma + \omega^{(1)}) + B + 2\omega_0 a_{13}(\omega^{(1)}\eta)' = 0 \tag{22}$$

whose solution is

$$\omega^{(1)} = -\sigma \tag{23}$$

In this case steady state solutions correspond to the solution of the cubic equation. For such detuning parameter that scales as $f^2$ the solution for $a_{13}$ is of order one. Thus, once again, small perturbations in the stellar gravitational potential give rise to finite response of the gas flow.

Figure 1 presents the resonant solutions in the velocity plane and gas density profiles compared with non-resonant solutions.

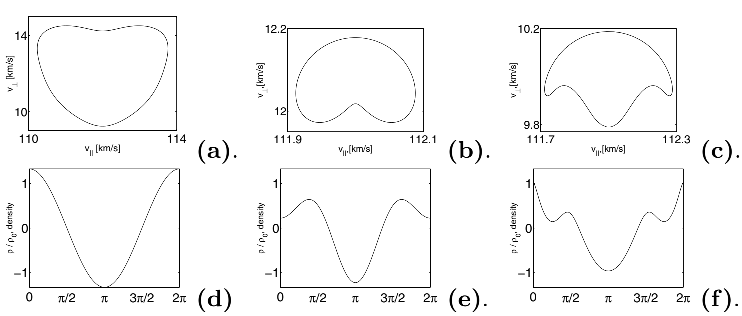



**Fig. 1.** Flow in velocity plane and normalized gas density profiles. Three cases are presented: 1. **(a)** and **(d)** - without resonance for distance from galactic center, r=10kpc, 2. **(b)** and **(e)** - resonance one-two for r=12.5 kpc, 3. **(c)** and **(f)** - resonance one-three for r=13.5 kpc. The variation of the normalized density $\rho/\rho_0$ presented as a function of the phase angle $\eta$.

## 4 Interpretation of the Results

The importance of corotation and Lindblad resonances has been recognized long ago see [22–25]. Particularly, it was found by calculation of the orbits that the main families of orbits in realistic model of spiral galaxy enhance the spiral arms up to the resonance $\chi/(\Omega - \Omega_p) = 4/1$. In this paper we have employed the hydrodynamic approach following [11] and have seen that the ultra harmonic resonances exist if the natural frequency $\omega_0$ is a rational multiple of the forcing frequency. In our case the natural frequency is given by

$$\omega_0 \approx \frac{\chi}{2(\Omega - \Omega_p)}(1 + \frac{a^2k^2}{8(\Omega - \Omega_p)^2 \cos^2 i})$$

and forcing frequency is equal to unity. So considered above ultra harmonic resonances $\omega_0 \approx 2$ and $\omega_0 \approx 3$ correspond to 4/1 and 6/1 respectively.

The substantially different spatial arrangement of spirals of young and old objects would appear to be the most remarkable features [15]. This circumstance should be an important factor in the theory of the origin of the spiral structure, in attempts to explain the observational data. Therefore we suppose that non-linear gas response effects on spiral density wave that is created by old objects could be responsible for the appearance of structures similar to the observed ones. To show this we have plotted a chart (Fig. 2) as the variation of the gas density obtained in our calculations. We see from Fig. 2 that the secondary compression associated with the resonance one-two obtained in our calculations for distance from galactic center 10.5 kpc up to 11.5 kpc that produces the arm bifurcation and may well account for major spiral features. Notice that a bifurcation of gaseous spiral arms was modeled numerically in [27] and was associated also with the presence of nonlinear effects at the 4/1 ultraharmonic

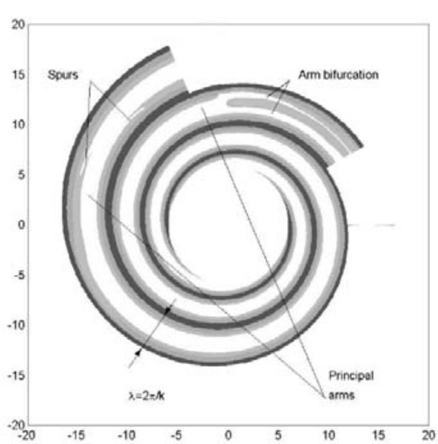



**Fig. 2.** Distribution of the gas density in response to galactic spiral density wave for two-arms mode and $\Omega_p = 13.5$ km/sec/kpc. The darkness of the chart is proportional to the value of the density (The black line shows the minimum)

resonance. The resonance one-three is found in our calculations to extend in distance from galactic center 13.5kpc up 14 kpc produce relatively short spurs and not major spiral features.

The authors thanks Yuriy N. Efremov and Arthur D. Chernin from the Sternberg Astronomical Institute, Moscow State University for their interest in this work. The authors have benefited from helpful discussions with Michael Gedalin and Evgeny Griv from the Ben-Gurion University of the Negev in Israel. Also, the authors are grateful to an anonymous referee of the paper for his/her useful comments on the original manuscript.

## References


1. C.C. Lin, C. Yuan, and F.H. Shu: Astrophys. J. **155**, 721 (1969)
2. W.W. Roberts: Astrophys. J. **158**, 123 (1969)
3. S.B. Pikel'ner: Astron. Zh. **47**, 752 (1970)
4. P.R. Woodward: Astrophys. J. **195**, 61 (1975)
5. M. Fujimoto: 'Modeling of gas flow through a spiral sleeve'. In: *Non-stable Phenomena in Galaxies. IAU Symp. No. 29, Byurakan, May 4-12, 1966.* (The Pubblishing House of the Academy of Sciences of Armenian SSR Yerevan 1968) pp.453-463
6. B.T. Lynds: 'The Distribution of Dark Nebulae in Late-Type Spirals'. In: *The Spiral Structure of our Galaxy, IAU Symp. No.* 38, ed. W.Becker and G.Contopoulos, (Reidel, Dordrecht 1970) pp.26-34
7. D.S. Mathewson, P.S. van der Kruit, and W.N. Brown: Astron. Astrophys., **17**, 468 (1972)
8. Y. Sofue: Publ. Astron. Soc. Japan, **37**, 507 (1985)
9. V.G. Berman, and Yu.N. Mishurov: Astrofisics, **16**, 52 (1980)
10. F.H. Shu, V. Milione, W. Gedel, C. Yuan, D.W. Goldsmith, and W.W. Roberts: Astrophys. J. **173**, 557 (1972)
11. F.H. Shu, V. Milione, and W.W. Roberts: Astrophys. J. **183**, 819 (1973)
12. S.A. Kaplan, S.B. Pikel'ner: Annual review of astronomy and astrophysics **12**, 113 (1974)
13. L.S. Marochnik, A.A. Suchkov: Soviet Physics - Uspekhi **17**, 85 (1974)
14. K. Rohlfs: *Lectures on Density-Wave Theory.* Lect. Notes. Phys. No. 69, (Springer, Berlin 1977)
15. F.Zwicky: *Morphological Astronomy* (Springer-Verlag, Berlin 1957)
16. B.A. Vorontsov-Vel'iaminov: *Extragalactic astronomy* (Harwood Academic, Chur, Switzerland 1987)
17. D.L. Block, R.J. Wainscoat: Nature **353**, Sept. 5, 48 (1991)
18. A.N. Nelson, and T. Matsuda: Month. Not. RAS, **179**, 663 (1977)
19. W.W. Roberts, Jr.: PASP, **105**, 670 (1993)
20. E. Liverts: Astrophys. & Space Sci. **274**, 513 (2000)
21. L.D. Landau, and E.M. Lifshitz: *Mecanics* (Pergamon Press, Oxford 1960)
22. B. Lindblad:, Handbuch der Physik, **53**, 21 (1959)
23. D. Lynden-Bell, and A.J. Kalnajs: Month. Not. RAS, **157**, 1 (1972)
24. G. Contopoulos, and P.Grosbøl: Astron. Astrophys., **155**, 11 (1986)
25. R.J. Allen, B. Canzian, and S.H. Lubow: PASP, **105**, 664 (1993)
26. P. Artymowicz, and S.H. Lubow: Astrophys. J., **389**, 129 (1992)
27. P.A. Patsis, P.Grosbøl, and N. Hiotelis: Astron. Astrophys., **323**, 762, (1997)

# The Level of Chaos in N-Body Models of Elliptical Galaxies

Nikos Voglis[1], Costas Kalapotharakos[1,2], Iioannis Stavropoulos[1,2], and Christos Efthymiopoulos[1]

[1] Academy of Athens, Research Center for Astronomy
[2] University of Athens, Department of Physics, Section of Astrophysics

**Abstract.** This paper presents a combination of methods that estimate reliably the level of chaos (proportion of particles in chaotic orbits and the corresponding values of the Lyapunov Characteristic Numbers) in self-consistent N-body models of elliptical galaxies. A careful simultaneous use of several numerical tools can induce the proportion of particles in chaotic orbits dynamically important within one Hubble time. In models with smooth centers the mass component in chaotic motion is less than about 30% of the total mass. In models with central black holes this percentage increases up to 70%. Typical Lyapunov characteristic numbers are below 0.1 in units of the inverse crossing time. A remarkable property of the chaotic mass component is that it has a different surface density profile than that of the ordered component. The superposition of the two profiles causes observable humps in the overall profile, which are suggested as a possible observational 'signature' of chaos in elliptical galaxies.

## 1 Introduction

An important open problem of stellar dynamics is the level of chaos in realistic self-consistent stellar systems. In this paper we study the level of chaos in four different self-consistent N-Body models of elliptical galaxies in equilibrium.

Two of the N-Body systems (Q and C models) are produced from quiet and clumpy cosmological initial conditions respectively [6], [4]. These models are non-rotating and they have a smooth density profile at the center. The other two models (QB1 and QB2) are produced from the Q model by adding a point mass (black hole) at the center, with a mass equal to 0.1% and 1% of the total galactic mass respectively. All the models are triaxial, but the Q and QB models are more elliptical than the C-models.

The self-consistent potential at equilibrium is realized by the N-Body code [1] as a smooth series of a radial plus spherical harmonic expansion. Near equilibrium, the expansion coefficients have almost constant values. Then the system is represented approximately by a 3D autonomous Hamiltonian

$$H = \frac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2) + V(x, y, z) \tag{1}$$

The equilibrium configurations are triaxial in all models, taken as $x$ the direction of the shortest axis and $z$ the direction of the longest axis.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 117–125, 2003.

## 2 Method to Distinguish Regular and Chaotic Orbits

The most chaotic orbits are identified by defining a particular threshold in the values of their Lyapunov Characteristic Number ($LCN$). Orbits below this threshold, although chaotic, behave macroscopically like regular orbits within a Hubble time.

The method, used to distinguish between regular and chaotic orbits, faces the problem of the very different (by two orders of magnitude) orbital periods of particular orbits within the same system. When one calculates Lyapunov times, must decide whether these times should be expressed in terms of the particular periods of the orbits, or of the half mass crossing time (average dynamical period) in the system. The present method provides a compromise for this problem and gives a not very much biased estimator of the chaoticity of the orbits, as the latter reflects to the macroscopic properties of the system within a Hubble time. The method uses three different indicators.

### 2.1 Specific Finite Time Lyapunov Characteristic Number

The Specific Finite Time Lyapunov Characteristic Number $L_j$ for the orbit of the particle $j$ is given by the formula

$$L_j(T_{rj}, t_j) = \frac{T_{rj}}{t_j} \sum_{i=1}^{N_j} a_{ij} \tag{2}$$

where $T_{rj}$ is the average radial period of the orbit, $t_j$ is the integration time, $N_j$ is the number of time steps $\Delta t = t_j/N_j$ and $a_{ij}$ is the stretching number [7] at the time step $i$, ($i = 1, ...N_j$). The stretching number $a_{ij}$ in (2) is defined in terms of the length of the deviation vector $\xi_j(t_i)$ from the orbit $j$ at the time $t_i$ in the six-dimensional phase space, by the equation

$$a_{ij} = \ln \frac{\xi_j(t_i + \Delta t)}{\xi_j(t_i)} \tag{3}$$

The deviation vectors $\xi_j(t_i)$ are calculated by numerical integration of the variational equations of motion.

The Hubble time is taken equal to 100 half mass crossing times ($T_{hmct}$) of the system. The values of $L_j$ of all the orbits are calculated for the same number of radial periods $t_j/T_{rj} = 1200$. Thus, the particles of even the shortest radial periods ($T_{rj} \geq T_{hmct}/300$) are integrated for more than 4 Hubble times. This integration time detects chaotic orbits with $L_j$ stabilized at values no smaller than $10^{-3}$. This threshold is rather arbitrary but it is satisfactory for all practical purposes.

### 2.2 Common Unit Finite Time LCN

The $L_j$s are converted in common units by defining the Common Unit Finite Time Lyapunov Characteristic Number $L_{cu}$, as $L_{cu} = L_j T_{hmct}/T_{rj}$, in units of $1/T_{hmct}$. This number compares the chaotic orbits as regards their efficiency within a Hubble time.

### 2.3 Alignment Index

The method of Alignment Index [5] is based on some known properties of the time evolution of deviation vectors [8], [9]. Consider the evolution of two arbitrary different initial deviation vectors $\xi_{j1}$ and $\xi_{j2}$ of the same orbit. If an orbit is chaotic, the two deviation vectors tend to become parallel or anti-parallel to each other (depending on the initial values of the deviation vectors). If the orbit is regular, the two vectors tend to become tangent to the surface of an invariant torus and oscillate with respect to each other. This difference is measured by taking the minimum of the quantities $d_{j-}(t) = |\boldsymbol{\xi}_{j1}(t) - \boldsymbol{\xi}_{j2}(t)|$ (parallel deviation vectors) and $d_{j+}(t) = |\boldsymbol{\xi}_{j1}(t) + \boldsymbol{\xi}_{j2}(t)|$ (anti-parallel deviation vectors). This is called the Smaller ALignment Index, (SALI), or simply Alignment Index, (AI). For chaotic orbits, it reaches the limit of the computer accuracy ($\approx 10^{-16}$) at the end of the integration time. For regular orbits it is improbable to be less than $10^{-3}$ all along the integration time. Thus, the distinction made by the AI method is clear and fast, even for very weakly chaotic orbits.

## 3 Results

Due to computing time limitations, it is only possible to calculate the orbits of a representative sample of particles of the whole system, namely, one in every four particles uniformly distributed along the whole set of particles. Using as initial conditions the three coordinates and velocities of each particle of the sample, each orbit is integrated in the Hamiltonian (1) for a maximum time $t_j = 1200 T_{rj}$.

In Fig. 1a the values of the $L_j$ of the particles in the sample of the Q-model are plotted against their Alignment Indices at $t_j/T_{rj} = 20$, in log-log scale. Most points appear concentrated in a single group of triangular shape around a mean value of $Log(L_j) \simeq -1.4$ with $Log(AI) > -3$. This is called a **regular group**. A number of points form a lane emanating from the upper end of the regular group towards smaller values of AI. The points on the lane correspond to orbits that have just started indicating their chaotic character.

At $t_j/T_{rj} = 100$ (Fig. 1b) the main part of the regular group is displaced towards lower values of $L_j$ following a $t^{-1}$ law. However, a good number of points have followed a streaming motion along the lane towards smaller values of AI and larger values of $L_j$ and tend to form a **chaotic group**.

As time increases the number of orbits in the chaotic group increases, but more and more slowly. At $t_j/T_{rj} = 1200$ (Fig. 1c) the chaotic group is well separated from the regular group. The streaming of points along the lane becomes slower, but non-negligible. The points on this lane correspond to weakly chaotic orbits.

By introducing a threshold $L_j = 10^{-2.8}$ as the value separating the chaotic from the regular group, and a threshold $AI = 10^{-3}$, we find that the fraction of the chaotic orbits (chaotic component) corresponds to about 32% of the total mass. The rest is called regular component. If the points along the transport lane

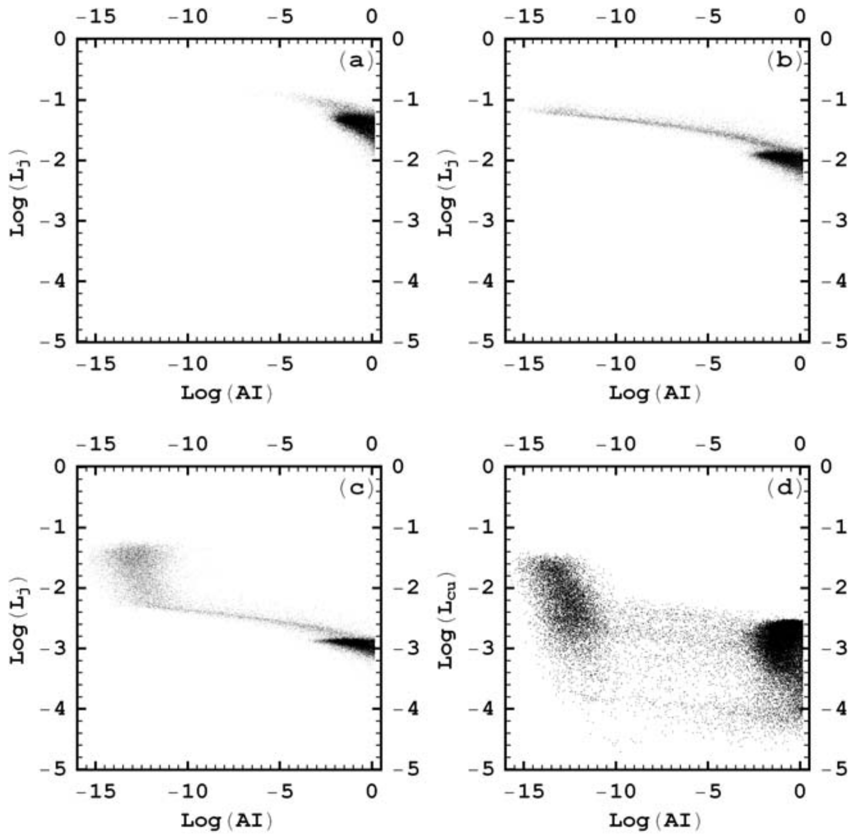



**Fig. 1.** The evolution of the orbits of the particles in the sample on the plane of the Log(AI)-Log(L) at **(a)** 20, **(b)** 100, **(c)** 1200 radial periods. In **(d)** we plot the Finite Time LCN in common units of $(1/T_{hmct})$, as derived from the data of (c)

are not counted with the chaotic orbits, then we find that the 'strictly chaotic' orbits ($AI \leq 10^{-10}$) are about 26%.

If $L_j$ is converted in units of the inverse half mass crossing time $T_{hmct}$ of the system, we find $L_{cu} = L_j T_{hmct}/T_{rj}$. Then, the results of Fig. 1c are converted to those shown in Fig. 1d. The $L_{cu}$ of the detected chaotic orbits range between $10^{-4.6}$ and $10^{-1.4}$, with a preference above the value of $10^{-3}$. It is obvious that a number of orbits that have been characterized as chaotic (with $L_j \geq 10^{-2.8}$) have values of $L_{cu}$s much smaller than the minimum value of $L_j$, because of their long radial periods. These small values of the $L_{cu}$, describe the very small diffusion in a Hubble time.

Figure 2 shows the same planes $L_j$ vs. $AI$ and $L_{cu}$ vs. $AI$ as in the Q model (Fig. 1c,d) for the experiments C (Fig. 2a,b), QB1 (Fig. 2c,d) and QB2 (Fig. 2e,f). The C model contains a smaller number of chaotic orbits and a somewhat smaller appearing maximum value of $L_j$ and $L_{cu}$. On the other hand, chaos is much more abundant if we add a black hole (Fig. 2c,d model QB1 with a

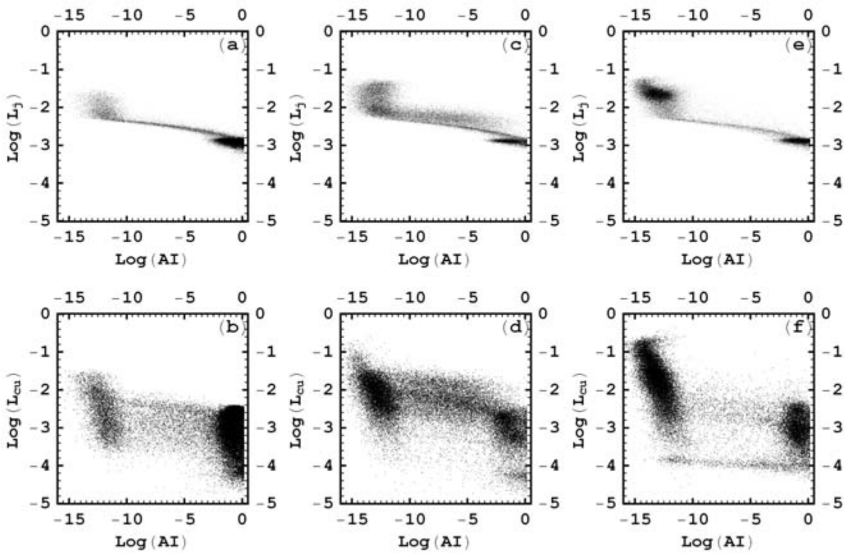



**Fig. 2.** **(a,b)**,**(c,d)**,**(e,f)** as Figs. 1c,d but for C, QB1, QB2 model respectively. Notice that the QB models (especially QB2) have orbits with $L_{cu}s$ above $10^{-1}$

.

black hole of 0.1% and Fig. 2e,f. model QB2 with a black hole of 1% of the total galactic mass). As the mass of black hole increases, there are more particles in the region of higher Lyapunov numbers ($L_j$ and $L_{cu}$). However, even in the QB2 model $L_j$s remain well below the value of $10^{-1}$. A small fraction of particles have $L_{cu}$ larger than $10^{-1}$ due to their very short radial periods. There is a pronounced transport of particles from regions of low $L_j$ (or $L_{cu}$) to regions of larger $L_j$ in the QB experiments.

Figure 3 shows the **time evolution of the proportion of chaotic orbits** for all the systems. Fig. 3a shows the time evolution of the percentage of the particles in orbits characterized as chaotic in the Q model. The solid line corresponds to a strict criterion of chaoticity, i.e. $AI < 10^{-10}$. The dashed line corresponds to a more flexible criterion, including the particles on the transport lane. About 30% of the particles in the Q model are characterized as chaotic after 1200 half mass crossing times. A smaller percentage (about 25%) is found in the case of the C experiment (Fig. 3b). On the other hand, the percentage of chaotic orbits is much higher if we add a black hole (QB1 and QB2 models, Figs. 3c,d), being as high as 70% in the case of the model QB2.

An important effect of this large increase of the level of chaos in the QB experiments is in **the distribution of the energies** (number density function) of the various systems. The distribution of all the particles along the energy axis for the Q experiment is shown in Fig. 4a by a solid line. The dashed line gives the distribution of the particles of the detected chaotic component. In the region of small energies (i.e. below the energy level of about $-60$) no chaotic orbits were detected by this threshold of chaoticity. The majority of particles

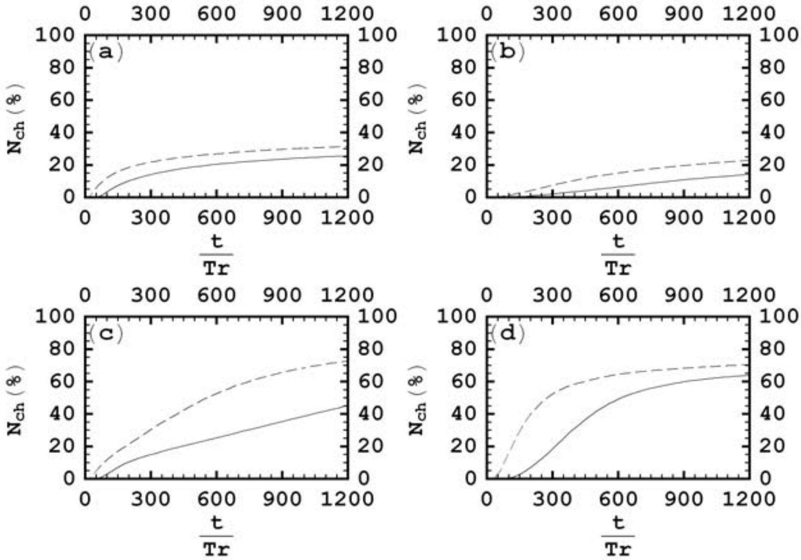



**Fig. 3.** The evolution of the percentage of orbits characterazed as chaotic in various models (**(a)** Q, **(b)** C, **(c)** QB1, **(d)** QB2). The solid line corresponds to the "strictly" chaotic part ($AI < 10^{-10}$) while the dashed line includes the orbits on the lane. The different rates of growth reflects the difference in the level of chaos

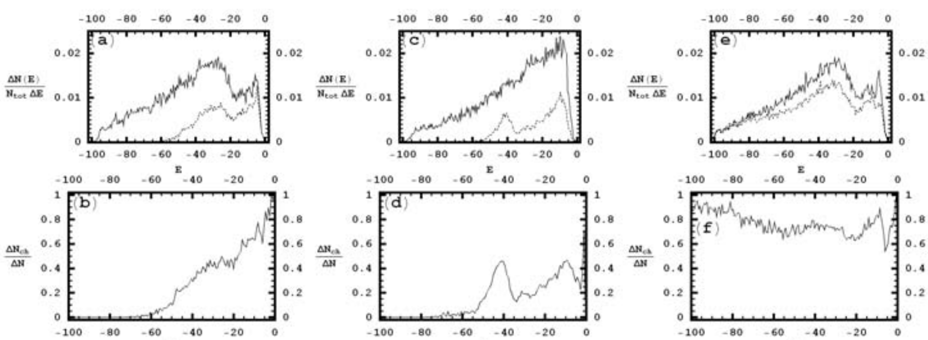



**Fig. 4. (a), (c), (e)** The distribution of the total mass (solid line) and of the detected chaotic part (dashed line) along the energy axis in the Q, C, QB1 model respectively. **(b), (d), (f)** The ratio of the number of chaotic orbits to the total number of orbits at every bin of energy along the energy axis in the Q, C, QB1 model respectively

in chaotic orbits with large binding energies spend most time at large radii and they mainly contribute in forming the halo of the galaxy. This effect is clearly seen in Fig. 4b, which gives the **relative ratio of particles in chaotic orbits** along the energy axis for the Q experiment.

The same distribution for the C model is seen in Figs. 4c,d. There is a remarkable difference between the Q and C experiments. In the Q model (Fig. 4b) the ratio of the detected chaotic orbits to the total number of orbits at every energy bin increases almost monotonically. In contrast, in the C model (Fig. 4d) this ratio has a pronounced maximum at the energy level of $\approx -40$ followed by a minimum around the energy level of $\approx -30$. This difference can be explained by noticing that in the C model there is a good number of ordered 1:1 tube orbits, at the energy levels around $-30$. These orbits, together with the chaotic orbits that have an almost spherical distribution, balance the effect of the box or box-like orbits, so that the asphericity of the system is reduced to the value required by the self-consistent equilibrium. On the other hand, in the Q model, at this energy levels, there is an extensive stable area in phase space corresponding to the 1:1 tube orbits, but it is almost empty, i.e. it is occupied by only a small number of orbits [2], [3]. At the same energy levels the Q model possesses a good number of chaotic orbits. These orbits are flexible to follow boxy or circular geometries but they are almost spherically distributed. Thus they prevent large departures of the system from sphericity.

If we add a black hole (Figs. 4e,f, model QB1), the chaotic orbits are no longer limited to small absolute binding energies but their distribution extends all the way to energies corresponding to the central value of the potential. This is because the black hole destroys the regular character of most box orbits close to the center, by causing large deflections of the orbits.

We finally compare the **surface densities** of the projections of the particles on various planes for the representative models Q and QB2. Fig. 5a,c shows the projections of the particles in ordered orbits for the Q experiment on the plane x-z and y-z respectively, while Fig. 5b,d shows the same projections but for the particles in chaotic motion. The main conclusion is that the large ellipticity of this galaxy is due to the regular orbits mainly, while the chaotic component tends to make the galaxy more spherical. On the other hand, the addition of a black hole (model QB2, Figs. 6a,b,c,d) has the effect of increasing the number of chaotic orbits. Thus the galaxy becomes more spherical in the presence of a black hole. Notice that the projection of regular particles on the plane y-z (Fig. 6c) is almost spherical. This is due to the fact that regular orbits move mostly at 1:1 tube orbits. For that reason the models with black hole tends to be more oblate.

The combination of the two surface density profiles (regular and chaotic) has the effect that the logarithmic slope ($s(r) = \frac{d \ln \sigma(r)}{d \ln r}$) of the overall profile $\sigma(r)$ forms an observable hump (Fig. 5e,f,g,h), [10], especially if the surface density profiles are taken along the shortest axis of the projection. Such a hump may be an observable signature of chaos in non-rotating elliptical galaxies.

## 4 Conclusions

We propose a methodology to obtain reliable estimates on the level of chaos in a self-consistent galactic system. This methodology combines three different numerical methods known in the literature. The combined use of the three methods provides a solution to the problem of estimation of Lyapunov times despite the very different periods of particular orbits within a galactic system.

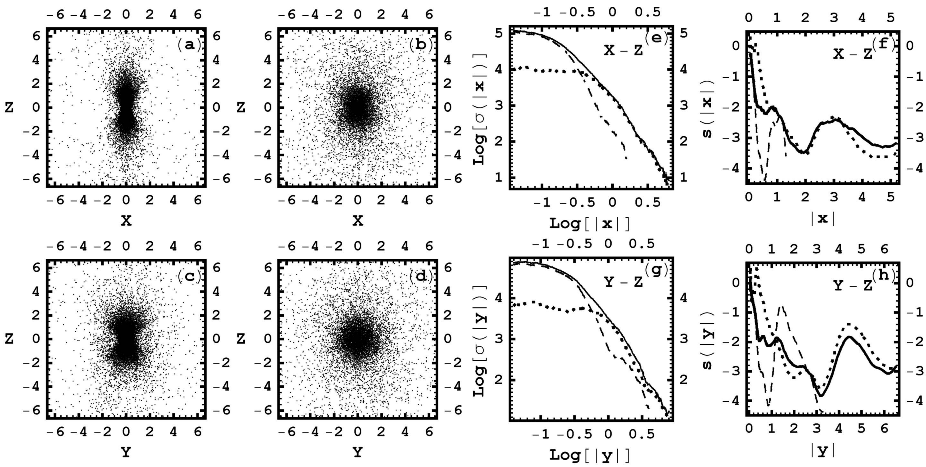



**Fig. 5.** The Q model. Projection on the x-z plane of the particles in ordered motion **(a)** and of the particles in chaotic motion **(b)**. In **(c)** and **(d)** as in (a) and (b), but on the y-z plane. In **(e)**, **(g)** is shown the surface density profiles along a slit on the short axis of each projection in Log-Log scale. In **(f)**, **(h)** is shown the Logarithmic slope of the surface density as a function of the distance along the same slit. Solid lines refer to the total mass, the dashed lines to the mass in ordered motion and the dotted lines to the mass in chaotic motion, The different slopes of the two components create a hump at about the half mass radius of the system

The models Q and C (with smooth central density profiles) have chaotic orbits only at relatively low absolute energies, i.e. at energy levels exceeding the deepest 30% of the potential well. Below this level most orbits are regular boxes or box-like. In the Q model, the detected chaotic part is about 30% of the total mass. This part has a nearly spherical distribution. It imposes limitations on the maximum ellipticity of the system, despite the fact that only a part less than about 8% of the total mass moves in chaotic orbits able to develop chaotic diffusion within a Hubble time. In the C model, the detected chaotic part is about 25% of the total mass, but only less than 2% can develop chaotic diffusion within a Hubble time.

Chaos is much more pronounced in the QB models with central black holes, and it extends to energies reaching the minimum of the potential well. This has implications on the number of particles in box or 1:1 tube orbits, and it affects the ellipticity of the systems. The overall proportion of particles in chaotic orbits reaches as much as 70% in the QB models. The more massive central black hole model (QB2) produces chaotic orbits with higher values of $L_{cu}s$, but the limit of $10^{-1}$ is hardly exceeded.

In all the systems, the chaotic components produce different surface density profiles than these of the rest of the mass. The combination of the two profiles produces observable signatures of chaos in non-rotating elliptical galaxies.

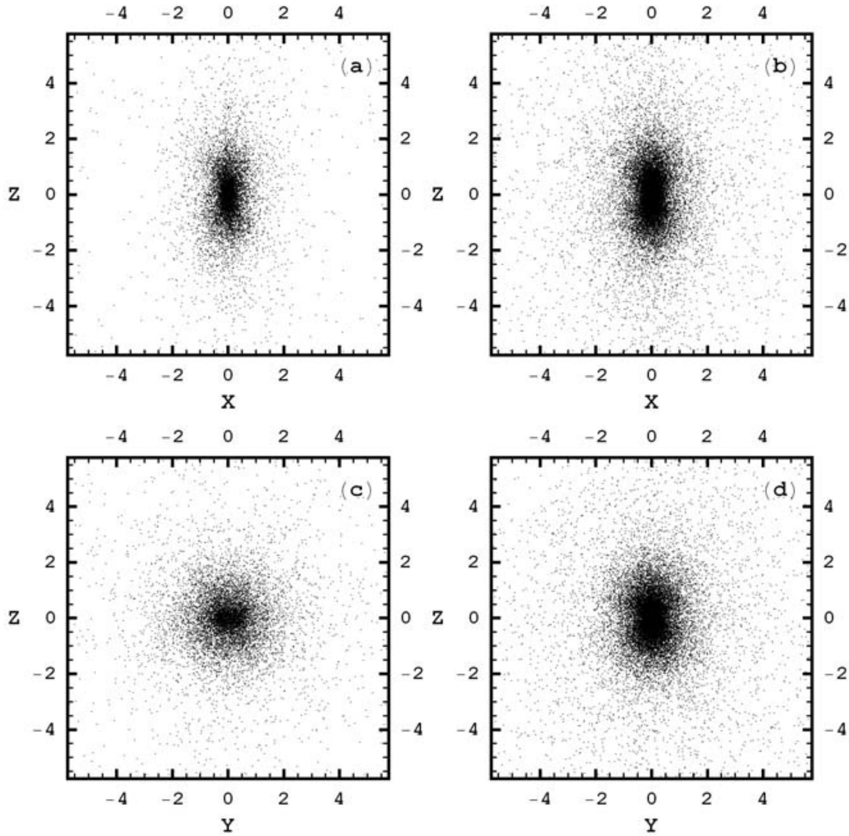



**Fig. 6.** As in Fig. 5, but for the QB2 model. We see that the projection of the particles in ordered motion forms a nearly spherical distribution because in that case the ordered motions are mostly 1:1 tube

## References


1. A. J. Allen, P. L. Palmer, J. Papaloizou: MNRAS **242**, 576 (1990)
2. G. Contopoulos, C. Efthymiopoulos, N. Voglis: CeMDA **78**, 243 (2000)
3. G. Contopoulos, N. Voglis, C. Kalapotharakos: CeMDA **83**, 191 (2002)
4. C. Efthymiopoulos, N. Voglis: A&A **378**, 679 (2002)
5. Ch. Skokos: JPhA **34**, 10029 (2001)
6. N. Voglis: MNRAS **267**, 379 (1994)
7. N. Voglis, G. Contopoulos: JPhA **27**, 4899 (1994)
8. N. Voglis, G. Contopoulos, C. Efthymiopoulos: PhRvE **57**, 372 (1998)
9. N. Voglis, G. Contopoulos, C. Efthymiopoulos: CeMDA **73**, 211 (1999)
10. N. Voglis, C. Kalapotharakos, I. Stavropoulos: MNRAS **337**, 619 (2002)

# Low Frequency Power Spectra and Classification of Hamiltonian Trajectories

George Voyatzis

Department of Physics, University of Thessaloniki, 54124 Thessaloniki, Greece

**Abstract.** We consider the problem of trajectory classification (as regular or chaotic) in Hamiltonian systems through power spectrum analysis. We focus our attention on the low frequency domain and we study the asymptotic behavior of the power spectrum when the frequencies tend to zero. A low frequency power estimator $\gamma$ is derived that indicates the significance of the relative power included by the low frequencies and we show that it is related to the underlying dynamics of the trajectories. The asymptotic behavior of $\gamma$ along a trajectory is qualitatively similar to that of the finite time Liapunov characteristic number. The standard map is used as a test model, because it is a typical model for describing Hamiltonian dynamics.

## 1 Introduction

Considering a variable $x(t)$ along a trajectory of a dynamical system, its (right side) power spectrum, defined as

$$p(f) = \lim_{T\to\infty} \frac{1}{T} \left| \int_0^T x(t) e^{-i2\pi f t} dt \right|^2, f > 0, \tag{1}$$

yields valuable information about the underlying local dynamics of the system. Particularly, Hamiltonian systems, having a $2n$-dimensional compact phase space, exhibit mainly two different types of dynamics, regular and chaotic, which are associated with qualitatively different power spectra [8, 10, 11]. Regular trajectories are wound on invariant tori, they are quasiperiodic and their power spectra are discrete (in the sense that they are described by few and well separated spectral peaks). Chaotic trajectories have spectra with more or less "grassy" background that indicates the existence of continuous spectral components. The above property is of significant importance in semiclassical dynamics [7]. Another important property of regular spectra is their invariant character, i.e. spectral peaks should be located at constant positions for different trajectory time segments and the fundamental frequencies $\omega_i, i = 1, .., n$ of a particular torus can be identified. But instead, chaotic spectra may show substantial changes through consequential segments indicating the non-existence of a torus. This characteristic has been proven a very useful tool in the study of long term trajectories in celestial mechanics [23].

From a theoretical point of view, quasiperiodicity implies spectra that are composed of lines at frequencies $f = \sum_{i=1}^{n} m_i\omega_i$, $m_i \in \mathbf{Z}$, i.e. the frequencies,

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 126–134, 2003.

where spectral peaks can appear, form a dense set on the real frequency axis of the spectrum. In general and under the absence of analytical solutions, trajectories are presented by sampled data for finite time intervals. Therefore, on the basis of numerical evidence, the distinction between regular and chaotic spectra becomes quite difficult when it refers to strongly deformed tori or weak chaotic motion. In [2] is shown that the above distinction, via finite time spectra, is impossible in principle and the true dynamics is revealed when $t \to \infty$.

Concerning the above intrinsic limitation, it is natural to think about asymptotic properties of the spectra for long time trajectory evolution. By increasing time, lower frequency modes are revealed in the spectrum. Then, the low frequency spectrum domain should be expected to provide a serious indication for characterizing the underlying dynamics. This paper attempts to handle the low frequency properties of power spectra obtained for bounded trajectories of Hamiltonian systems. Instead of calculating the whole power spectrum, the time evolution of a dynamical quantity $\gamma(t)$ along a trajectory is examined that reveals efficiently the requested information.

## 2 The Low Frequency Domain and Underlying Dynamics

For dynamical variables $x(t)$ that are well-behaving functions (i.e. they evolve smoothly in time, are bounded and have no singularities) their power spectra converge exponentially to zero as $f \to \infty$. Thus, we may define a high-frequency cut-off $f_H$ such that $p(f) \approx 0$ for $f > f_H$ [10]. This is the case independently on whether the trajectories are regular or chaotic. In a non strict way, the low frequency domain is defined as the frequency interval $\mathcal{L} = (0, f_0)$, where $f_0 \ll f_H$. When a low frequency cut-off $f_L \in \mathcal{L}$ can be defined, such that $p(f)$ converges to zero when $f \to 0$, then the power spectrum is called *convergent*, otherwise, is called *divergent*. There are indications that these two types of spectra are related to the type of the underlying dynamics. In [6] it is shown, through reordered spectra, that the quasiperiodic regime of circle maps corresponds to convergent power spectra ($p(f) \sim f$) while in [4] an exponential convergence is indicated. In [12] the amplitude of peaks, located at low frequencies, are associated with the effect of small denominators in the convergence of the classical perturbation series of near integrable systems. The convergence of these series, which implies that a torus persists the perturbation, implies also convergent spectra. In [1] it is mentioned that chaotic trajectories contains a “central peak” in $\mathcal{L}(f_L)$, which is a peak with considerable amplitude. Generally, by considering more precise computations, the “central peak” is proven to be an erratic continuous portion in the low frequency domain. It’s presence has been ascribed to the existence of a nearby separatrix trajectory [8, 10]. Additionally, chaotic trajectories may show $1/f^a$ power spectra ($a \approx 1$) even for Hamiltonian systems of few degrees of freedom. Such behavior suggest the existence of a slow diffusion process [3].

In Fig. 1 the rich dynamics of the standard map

$$x_{x+1} = x_n + k\sin(x_n + y_n) \ , \ y_{n+1} = x_n + y_n \ \ (mod\, 2\pi), \tag{2}$$

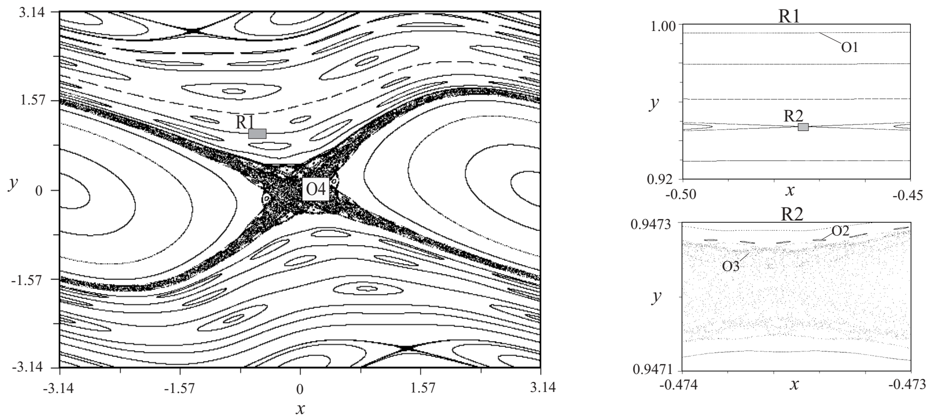



**Fig. 1.** The standard map dynamics for $k$ =0.7. Magnifications of the regions R1 and R2 are shown and the orbits, referred to spectra of Fig.2, are indicated by O1,O2,O3 and O4.

is presented on the plane $x - y$ for $k = 0.7$. The trajectory O1 is a typical quasiperiodic trajectory that evolves relatively far from a significant resonance. The trajectory O2 forms small islands and is located close to the narrow chaotic zone of the weak resonance 2:9 (region R2 of Fig. 1). The main chaotic region (O4) is obtained around the unstable fixed point at (0,0). The power spectra of these trajectories are shown in Fig. 2. Log-Log scales are used in order to emphasize the low frequency domain. For quasiperiodic trajectories, located far from significant resonances, discrete peaks constitutes the spectrum (Fig. 2a). Furthermore, we obtain that the peaks show a rapid decay in amplitude as $f \to 0$. When a quasiperiodic trajectory evolves on a torus close to a separatrix manifold, much more peaks are present both at high and low frequency domains (Fig. 2b). A typical characteristic in this case is the appearance of a family of peaks in some low frequency domain (e.g. the peaks surrounded by the dotted-line in Fig. 2b). However, the spectrum is convergent and this is always the case as long as the underlying dynamics is regular. Figure 2c refers to a trajectory that evolves inside the homoclinic web of the separatrix at the resonance 2:9 of width $\Delta y \approx 10^{-4}$. The low frequency domain shows a continuous distribution which approximates a Lorentzian spectrum shape [10]. We should note that this characteristic can not be noticed by using normal scales. Finally, Fig. 2d corresponds to the trajectory, evolving in the chaotic region around the unstable fixed point at (0,0). The spectrum seems to follow an $1/f^a (a \approx 0.8)$ divergence, but for this and all other trajectories of the standard map examined, the $1/f^a$-divergence is limited and the spectrum saturates, i.e. it tends to a constant nonzero value as $f \to 0$.

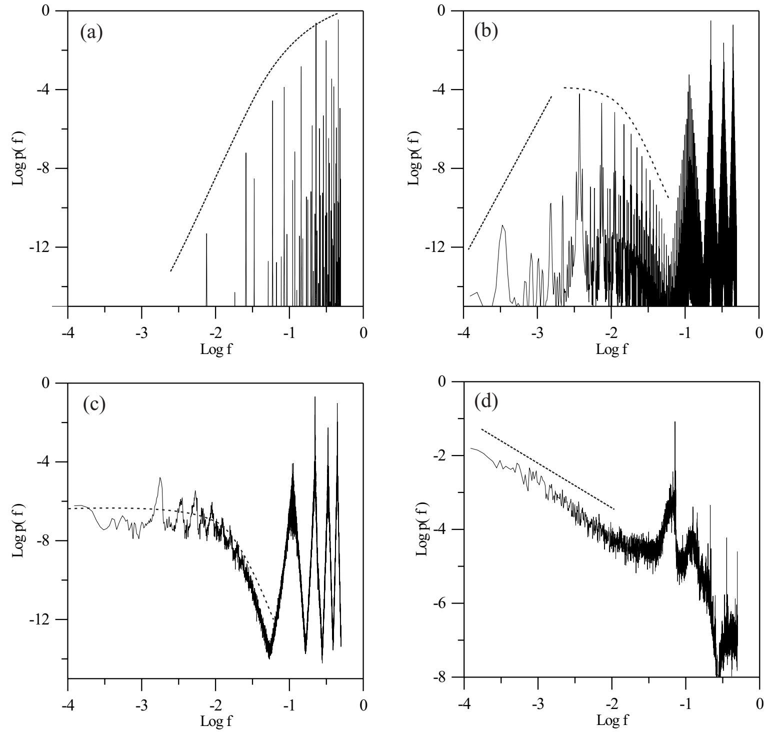



**Fig. 2.** Typical power spectra (a-d) for the orbits O1,O2,O3 and O4 respectively, shown in Fig.1.

## 3 Estimation of the Low Frequency Power and Asymptotic Characteristics

The asymptotic behavior of the power spectrum as $f \to 0$ can be estimated by examining the relative power included in $\mathcal{L}$. Considering a power spectrum $p = p(f)$, its normalized power spectral density at a frequency $f = f_0$ is defined as

$$\hat{p}(f_0) = \lim_{\Delta f \to 0} \frac{1}{\Delta f} \frac{I_p(f_0 - \Delta f/2, f_0 + \Delta f/2)}{I_p(0, \infty)}, \tag{3}$$

where we have used the notation

$$I_p(a, b) = \int_a^b p(f) df.$$

Since we are supposed to study the spectrum at low frequencies $f_0 \ll 1$, we may take $f_0 \in \mathcal{L}$ and $\Delta f = f_0$. Then, by taking into account the high frequency cut-off $f_H$, we approximate (3) by the quantity

$$\gamma^*(f_0) = \frac{1}{f_0} \frac{I(f_0/2, 3f_0/2)}{I(f_0, f_H)}. \tag{4}$$

We may consider $\gamma^*(f_0)$ to be a function of time along the particular trajectory by introducing the "period" $T = 1/f_0$,

$$\gamma(T) = \gamma^*(f_0) \ , \ \ f_0 \equiv 1/T. \tag{5}$$

In average, the power spectra of Hamiltonian trajectories follows profiles (smooth shapes), which can be described by the form

$$\tilde{p}(f) = \frac{p_0}{(\tau f)^a + \tau^2 f^2}, \ \ f > 0, \tau > 0, a \in (-a_0, a_0). \tag{6}$$

For high frequencies ($f \to \infty$), $p(f)$ decays rapidly to zero for any value of the parameter $a$, while $\tau$ is a frequency scaling factor [10]. For the low frequency domain we obtain convergent spectra for $a < 0$, the Lorentzian spectrum form for $a = 0$ and $1/f^a$-divergence for $a > 0$. It can be shown, that the corresponding $\gamma$ estimator, obeys the asymptotic behavior

$$\gamma(T) \propto T^a \ \text{ for } \ T \to \infty. \tag{7}$$

Thus, as $T \to \infty$, $\gamma(T)$ approximates the asymptotic behavior of the power spectrum when $f \to 0$.

Under numerical analysis, the variation of quantities along trajectories is given as time-series $x_k$. Namely, they are restricted in finite time intervals $[0, T]$ and are sampled at equidistant points $t_k = k\Delta t, k = 0, 1, ..., N-1$, where $N = [T/\Delta t]$. Then, by applying a discrete Fourier transform (DFT) [9], the right side power spectrum is approximated by the relation

$$p(f_n) = \frac{|H_n|^2}{N^2} \ , \ \ f_n = \frac{n}{N\Delta t}, \ n = 1, ..., N/2, \tag{8}$$

where $H_n$ denote the DFT coefficients. By setting $Dt = 1$, the calculated spectrum is restricted in the frequency domain $[f_1, f_{N/2}]$, where $f_1 = 1/N$ is the lowest available frequency and $f_{N/2} = 1/2$ is the high cutoff (Nyquist frequency). Then, we may let in (4) $f_0 = 1/N$, $f_H = 1/2$, $I_p(f_0/2, 3f_0/2) = p(f_1)$ and $I_p(f_0, f_H) = \sum_{n=1}^{n=N/2} p(f_k)$. Also, by taken into account the discrete form of the Parseval's theorem and writing $H_1$ in its trigonometric form, we obtain

$$\gamma(N) = \frac{\left(\sum\limits_{k=0}^{N-1} x_k \cos(2k\pi/N)\right)^2 + \left(\sum\limits_{k=0}^{N-1} x_k \sin(2k\pi/N)\right)^2}{\sum\limits_{k=0}^{N-1} x_k^2}, \tag{9}$$

where $N$ is the length of the trajectory sample that corresponds to integration time $T = N\Delta t$. The formula (9) does not allow the simultaneous calculation for both the trajectory and the estimator $\gamma$. However, $\gamma$ should be assumed as a dynamical variable along the trajectory.

## 4 Numerical Results and Conclusions

In Fig. 3 the evolution of $\gamma(N)$, compared to the finite Liapunov characteristic number $\lambda(N)$ is shown for the same trajectories as that of Fig. 2. For the calculation of $\gamma(N)$ we used the time series $z_n = \sin(x_n)$, $n = 0, ..., N-1$ in order to avoid the discontinuity caused by the modulo operation in (2). Cases (a) and (b) clearly indicate regular dynamics. Both, $\gamma(N)$ and $\lambda(N)$ converge rapidly to zero, almost as $1/N$. In case (c), which corresponds to a weakly chaotic trajectory, we obtain convergence, for both $\gamma(N)$ and $\lambda(N)$, which holds during that time the orbit is sticky. Afterwards, $\lambda(N)$ tends slowly to saturation, while $\gamma(N)$ shows an abrupt increment indicating that the trajectory entered the chaotic channel and reveals its chaotic nature efficiently. In case (d) $\lambda(N)$ seems to tend at a relatively large value indicating the chaotic character of the trajectory. The estimator $\gamma(N)$ shows a remarkable divergence indicating, additionally, a slow diffusion that is apparent at least up to the integration interval. Independently of the parameter $k$, cases (a) and (b) are typical for all regular trajectories, and, in average, it holds $\gamma \sim N^a$ with $a \approx 1$. It worths to note, that the convergent evolution of $\gamma(N)$ is followed by dense sharp peaks, towards to lower values. Such peaks indicate that the corresponding spectra should be

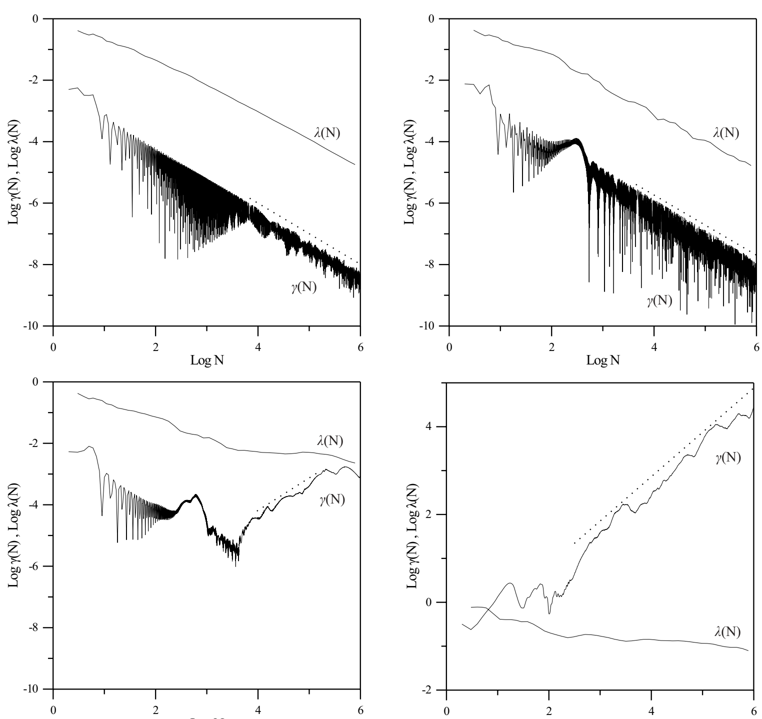



**Fig. 3.** The evolution of $\gamma(N)$ and the finite time Liapunov characteristic number for the trajectories referred in Fig. 2.

discrete as, indeed, they are. For chaotic trajectories, the evolution of $\gamma(N)$ is quite smooth because of the continuity of the associated spectra.

From Fig. 3 we observe that $\gamma(N)$ and $\lambda(N)$ show the same asymptotic characteristics that can be used to classify the trajectories as ordered or chaotic. Especially for ordered trajectories, $\gamma(N)$ and $\lambda(N)$ show a similar decay with respect to $N$. For chaotic orbits, $\gamma(N)$ and $\lambda(N)$ seem to reach saturation values in different ways, but nevertheless, we can conclude their non-convergent character. In Fig. 4 we plot the pairs $(\gamma(N), \lambda(N))$ obtained from 2000 randomly selected initial conditions for $k = 0.9$ (crosses) and $k = 1$ (circles) and after $N = 3 \cdot 10^6$ iterations. The 97% of the points belongs either in the domain A (regular trajectories) or in the domain B (chaotic trajectories). The rest of them, which mainly correspond to $k = 0.9$, may be classified as weakly chaotic trajectories. By increasing the number of iterations, the points in region A, and, generally, the points that correspond to ordered motion, move to lower values, since both $\gamma(N)$ and $\lambda(N)$ tend to zero. Therefore, the reliability of the trajectory classification increases as the length of the trajectory increases.

In many cases, $\gamma$ shows in average $N^a$-divergence, with $a \approx 1$. However, after a long time evolution, $\gamma$ shows small oscillations around a constant value $\bar{\gamma}$ (saturation). The value of $\bar{\gamma}$ is proportional to how long the $N^a$-divergence takes place and it is associated to the $1/f^a$-divergence of the spectrum caused by the slow diffusion through small islands and cantori [8]. Thus, large values of $\bar{\gamma}$ indicate the diffusive character of the trajectory rather than its strong chaotic evolution. In Fig. 5 we can see that the chaotic orbits, which start near the unstable fixed point (0,0) for $k \leq 1$, correspond to large values for $\bar{\gamma}$ (about $10^4$), but, as $k$ increases, we observe that $\bar{\gamma} \to 1$. According to the definition of $\gamma$, this value is obtained when the power spectrum has the form $p(f) \approx$const (white noise).

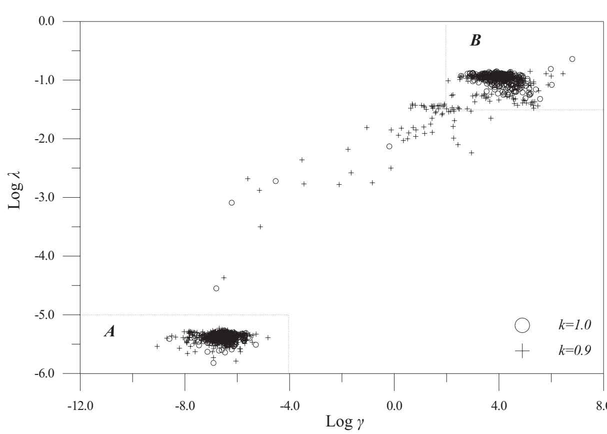



**Fig. 4.** The $(\gamma(N)$,$\lambda(N))$ graph for 2000 trajectories of the standard map. Points that belong to regions A or B indicates regular and chaotic dynamics, respectively, with great certainty.

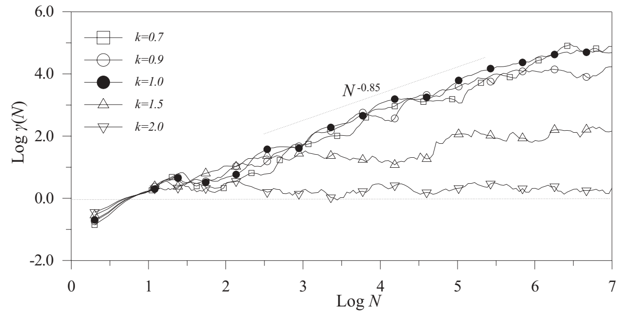



**Fig. 5.** The $\gamma(N)$ evolution along a chaotic trajectory close to the unstable fixed point (0,0) and for different values of the mapping parameter $k$.

Conclusively, we may claim that regular dynamics is associated with convergent spectra, in the sense that $p(f) \to 0$ as $f \to 0$ or, equivalently, $\gamma(T) \to 0$ as $T \to \infty$. Chaos is associated with divergent spectra where $p(f) \to$ const.$> 0$ as $f \to 0$ or, equivalently, $\gamma(T) \to$ const.$> 0$ as $T \to \infty$. In other words, for Hamiltonian bounded trajectories, discrete spectra are convergent, while, spectra with continuous local domains are divergent. Although there is not a rigorous proof, the numerical results support strongly the above relation.

The low frequency power estimator $\gamma$ can be used as an indicator for the qualitative character of a trajectory, in a similar manner to that of the Liapunov characteristic number $\lambda$. These two quantities are expressed in the same units (time$^{-1}$) but they have different physical meaning. Namely, $\gamma$ is associated with the long term regular or irregular evolution rather than the linear or exponential divergence of nearby trajectories and is calculated along a single trajectory.


## Acknowledgements

The author would like to thank Dr. S.Ichtiaroglou for the fruitful discussions about this topic.


## References


1. V.V.Beloshapkin, G.M.Zaslavsky: Phys. Lett. A **97**, 121 (1983)
2. R.S. Dumont, P. Brumer: J. Chem. Phys. **88**, 1481 (1988)
3. T. Geisel, A.Zacherl, G.Rabons: Phys. Rev. Lett. **59**, 2503 (1987)
4. S. Kim, S. Ostlund, G.Yu: Physica D **31**, 117 (1988)

23. J. Laskar: Physica D **67**, 257 (1993)

6. S.Ostlund, D.Rand, J. Sethna, E. Siggia: Physica D **8**, 303 (1983)
7. R.Roy, B.G.Sumpter, G.a.Pfeffer, S.K. Gray, D.W. Noid: Physics Rep. **205**, 111 (1991)
8. G.E. Powell: Regular and irregular frequency spectra. Ph.D.Thesis , Queen Mary College, London (1980)
9. T.R. Press, S.A. Teukolsky, W.T. Vetterling, B.P.Flannery: *Numerical Recipes in C*, (Cambridge University Press, Cambridge, 1992)

10. R.Z. Sagdeev, D.A. Usikov, G.M. Zaslavsky: *Nonlinear Physics* (Harwood Academic 1988)
11. M. Tabor: *Chaos and integrability in nonlinear dynamics*, (Wiley, New York 1989)
12. G. Voyatzis, S. Ichtiaroglou: J. Phys. A **25**, 5931 (1992)

# Low Frequency Power Spectra and Classification of Hamiltonian Trajectories

George Voyatzis

Department of Physics, University of Thessaloniki, 54124 Thessaloniki, Greece

**Abstract.** We consider the problem of trajectory classification (as regular or chaotic) in Hamiltonian systems through power spectrum analysis. We focus our attention on the low frequency domain and we study the asymptotic behavior of the power spectrum when the frequencies tend to zero. A low frequency power estimator $\gamma$ is derived that indicates the significance of the relative power included by the low frequencies and we show that it is related to the underlying dynamics of the trajectories. The asymptotic behavior of $\gamma$ along a trajectory is qualitatively similar to that of the finite time Liapunov characteristic number. The standard map is used as a test model, because it is a typical model for describing Hamiltonian dynamics.

## 1 Introduction

Considering a variable $x(t)$ along a trajectory of a dynamical system, its (right side) power spectrum, defined as

$$p(f) = \lim_{T\to\infty} \frac{1}{T} \left| \int_0^T x(t) e^{-i2\pi f t} dt \right|^2 , f > 0, \tag{1}$$

yields valuable information about the underlying local dynamics of the system. Particularly, Hamiltonian systems, having a $2n$-dimensional compact phase space, exhibit mainly two different types of dynamics, regular and chaotic, which are associated with qualitatively different power spectra [8, 10, 11]. Regular trajectories are wound on invariant tori, they are quasiperiodic and their power spectra are discrete (in the sense that they are described by few and well separated spectral peaks). Chaotic trajectories have spectra with more or less "grassy" background that indicates the existence of continuous spectral components. The above property is of significant importance in semiclassical dynamics [7]. Another important property of regular spectra is their invariant character, i.e. spectral peaks should be located at constant positions for different trajectory time segments and the fundamental frequencies $\omega_i, i = 1, .., n$ of a particular torus can be identified. But instead, chaotic spectra may show substantial changes through consequential segments indicating the non-existence of a torus. This characteristic has been proven a very useful tool in the study of long term trajectories in celestial mechanics [23].

From a theoretical point of view, quasiperiodicity implies spectra that are composed of lines at frequencies $f = \sum_{i=1}^{n} m_i\omega_i$, $m_i \in \mathbf{Z}$, i.e. the frequencies,

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 126–134, 2003.

where spectral peaks can appear, form a dense set on the real frequency axis of the spectrum. In general and under the absence of analytical solutions, trajectories are presented by sampled data for finite time intervals. Therefore, on the basis of numerical evidence, the distinction between regular and chaotic spectra becomes quite difficult when it refers to strongly deformed tori or weak chaotic motion. In [2] is shown that the above distinction, via finite time spectra, is impossible in principle and the true dynamics is revealed when $t \to \infty$.

Concerning the above intrinsic limitation, it is natural to think about asymptotic properties of the spectra for long time trajectory evolution. By increasing time, lower frequency modes are revealed in the spectrum. Then, the low frequency spectrum domain should be expected to provide a serious indication for characterizing the underlying dynamics. This paper attempts to handle the low frequency properties of power spectra obtained for bounded trajectories of Hamiltonian systems. Instead of calculating the whole power spectrum, the time evolution of a dynamical quantity $\gamma(t)$ along a trajectory is examined that reveals efficiently the requested information.

## 2 The Low Frequency Domain and Underlying Dynamics

For dynamical variables $x(t)$ that are well-behaving functions (i.e. they evolve smoothly in time, are bounded and have no singularities) their power spectra converge exponentially to zero as $f \to \infty$. Thus, we may define a high-frequency cut-off $f_H$ such that $p(f) \approx 0$ for $f > f_H$ [10]. This is the case independently on whether the trajectories are regular or chaotic. In a non strict way, the low frequency domain is defined as the frequency interval $\mathcal{L} = (0, f_0)$, where $f_0 \ll f_H$. When a low frequency cut-off $f_L \in \mathcal{L}$ can be defined, such that $p(f)$ converges to zero when $f \to 0$, then the power spectrum is called *convergent*, otherwise, is called *divergent*. There are indications that these two types of spectra are related to the type of the underlying dynamics. In [6] it is shown, through reordered spectra, that the quasiperiodic regime of circle maps corresponds to convergent power spectra ($p(f) \sim f$) while in [4] an exponential convergence is indicated. In [12] the amplitude of peaks, located at low frequencies, are associated with the effect of small denominators in the convergence of the classical perturbation series of near integrable systems. The convergence of these series, which implies that a torus persists the perturbation, implies also convergent spectra. In [1] it is mentioned that chaotic trajectories contains a "central peak" in $\mathcal{L}(f_L)$, which is a peak with considerable amplitude. Generally, by considering more precise computations, the "central peak" is proven to be an erratic continuous portion in the low frequency domain. It's presence has been ascribed to the existence of a nearby separatrix trajectory [8, 10]. Additionally, chaotic trajectories may show $1/f^a$ power spectra ($a \approx 1$) even for Hamiltonian systems of few degrees of freedom. Such behavior suggest the existence of a slow diffusion process [3].

In Fig. 1 the rich dynamics of the standard map

$$x_{x+1} = x_n + k \sin(x_n + y_n) \ , \ y_{n+1} = x_n + y_n \ \ (mod\, 2\pi), \tag{2}$$

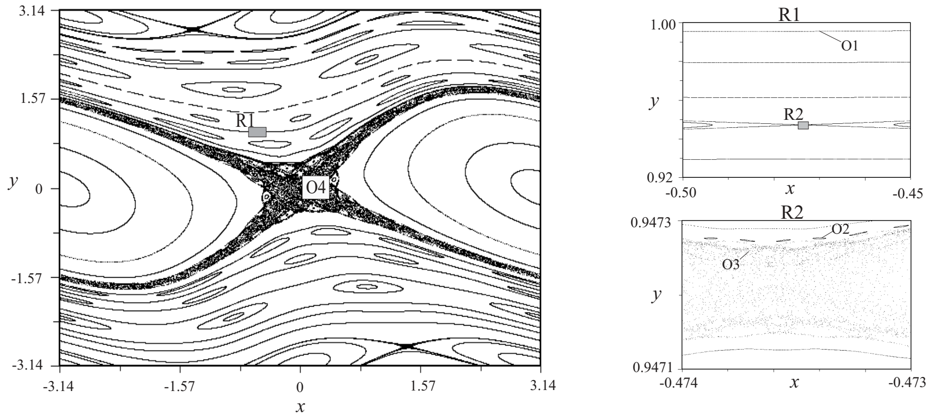



**Fig. 1.** The standard map dynamics for $k$ =0.7. Magnifications of the regions R1 and R2 are shown and the orbits, referred to spectra of Fig.2, are indicated by O1,O2,O3 and O4.

is presented on the plane $x - y$ for $k = 0.7$. The trajectory O1 is a typical quasiperiodic trajectory that evolves relatively far from a significant resonance. The trajectory O2 forms small islands and is located close to the narrow chaotic zone of the weak resonance 2:9 (region R2 of Fig. 1). The main chaotic region (O4) is obtained around the unstable fixed point at (0,0). The power spectra of these trajectories are shown in Fig. 2. Log-Log scales are used in order to emphasize the low frequency domain. For quasiperiodic trajectories, located far from significant resonances, discrete peaks constitutes the spectrum (Fig. 2a). Furthermore, we obtain that the peaks show a rapid decay in amplitude as $f \to 0$. When a quasiperiodic trajectory evolves on a torus close to a separatrix manifold, much more peaks are present both at high and low frequency domains (Fig. 2b). A typical characteristic in this case is the appearance of a family of peaks in some low frequency domain (e.g. the peaks surrounded by the dotted-line in Fig. 2b). However, the spectrum is convergent and this is always the case as long as the underlying dynamics is regular. Figure 2c refers to a trajectory that evolves inside the homoclinic web of the separatrix at the resonance 2:9 of width $\Delta y \approx 10^{-4}$. The low frequency domain shows a continuous distribution which approximates a Lorentzian spectrum shape [10]. We should note that this characteristic can not be noticed by using normal scales. Finally, Fig. 2d corresponds to the trajectory, evolving in the chaotic region around the unstable fixed point at (0,0). The spectrum seems to follow an $1/f^a (a \approx 0.8)$ divergence, but for this and all other trajectories of the standard map examined, the $1/f^a$-divergence is limited and the spectrum saturates, i.e. it tends to a constant nonzero value as $f \to 0$.

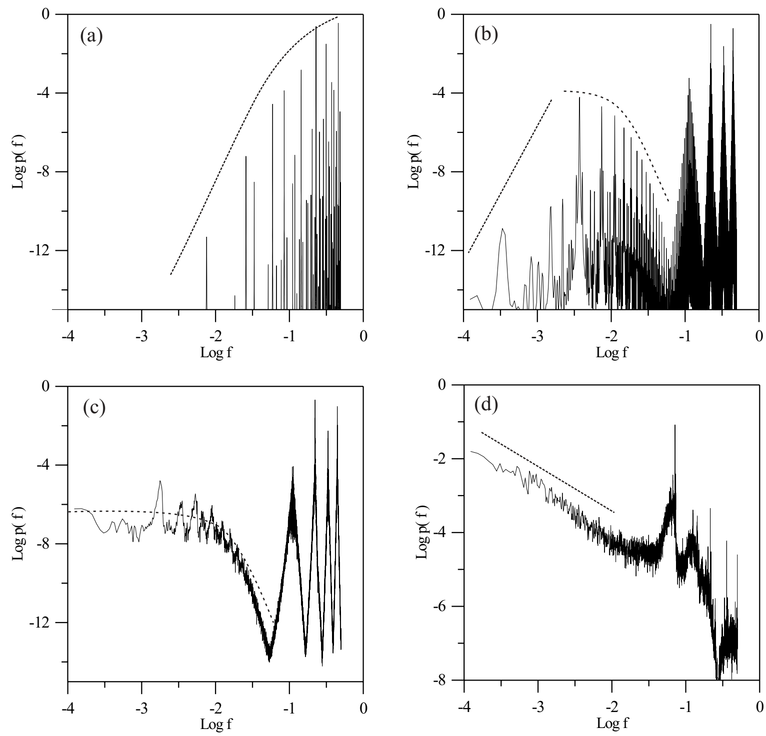



**Fig. 2.** Typical power spectra (a-d) for the orbits O1,O2,O3 and O4 respectively, shown in Fig.1.

## 3 Estimation of the Low Frequency Power and Asymptotic Characteristics

The asymptotic behavior of the power spectrum as $f \to 0$ can be estimated by examining the relative power included in $\mathcal{L}$. Considering a power spectrum $p = p(f)$, its normalized power spectral density at a frequency $f = f_0$ is defined as

$$\hat{p}(f_0) = \lim_{\Delta f \to 0} \frac{1}{\Delta f} \frac{I_p(f_0 - \Delta f/2, f_0 + \Delta f/2)}{I_p(0, \infty)}, \tag{3}$$

where we have used the notation

$$I_p(a,b) = \int_a^b p(f) df.$$

Since we are supposed to study the spectrum at low frequencies $f_0 \ll 1$, we may take $f_0 \in \mathcal{L}$ and $\Delta f = f_0$. Then, by taking into account the high frequency cut-off $f_H$, we approximate (3) by the quantity

$$\gamma^*(f_0) = \frac{1}{f_0} \frac{I(f_0/2, 3f_0/2)}{I(f_0, f_H)}. \tag{4}$$

We may consider $\gamma^*(f_0)$ to be a function of time along the particular trajectory by introducing the "period" $T = 1/f_0$,

$$\gamma(T) = \gamma^*(f_0) \ , \ f_0 \equiv 1/T. \tag{5}$$

In average, the power spectra of Hamiltonian trajectories follows profiles (smooth shapes), which can be described by the form

$$\tilde{p}(f) = \frac{p_0}{(\tau f)^a + \tau^2 f^2}, \ f > 0, \tau > 0, a \in (-a_0, a_0). \tag{6}$$

For high frequencies ($f \to \infty$), $p(f)$ decays rapidly to zero for any value of the parameter $a$, while $\tau$ is a frequency scaling factor [10]. For the low frequency domain we obtain convergent spectra for $a < 0$, the Lorentzian spectrum form for $a = 0$ and $1/f^a$-divergence for $a > 0$. It can be shown, that the corresponding $\gamma$ estimator, obeys the asymptotic behavior

$$\gamma(T) \propto T^a \text{ for } T \to \infty. \tag{7}$$

Thus, as $T \to \infty$, $\gamma(T)$ approximates the asymptotic behavior of the power spectrum when $f \to 0$.

Under numerical analysis, the variation of quantities along trajectories is given as time-series $x_k$. Namely, they are restricted in finite time intervals $[0, T]$ and are sampled at equidistant points $t_k = k\Delta t, k = 0, 1, ..., N-1$, where $N = [T/\Delta t]$. Then, by applying a discrete Fourier transform (DFT) [9], the right side power spectrum is approximated by the relation

$$p(f_n) = \frac{|H_n|^2}{N^2} \ , \ f_n = \frac{n}{N\Delta t}, \ n = 1, ..., N/2, \tag{8}$$

where $H_n$ denote the DFT coefficients. By setting $Dt = 1$, the calculated spectrum is restricted in the frequency domain $[f_1, f_{N/2}]$, where $f_1 = 1/N$ is the lowest available frequency and $f_{N/2} = 1/2$ is the high cutoff (Nyquist frequency). Then, we may let in (4) $f_0 = 1/N$, $f_H = 1/2$, $I_p(f_0/2, 3f_0/2) = p(f_1)$ and $I_p(f_0, f_H) = \sum_{n=1}^{n=N/2} p(f_k)$. Also, by taken into account the discrete form of the Parseval's theorem and writing $H_1$ in its trigonometric form, we obtain

$$\gamma(N) = \frac{\left(\sum_{k=0}^{N-1} x_k \cos(2k\pi/N)\right)^2 + \left(\sum_{k=0}^{N-1} x_k \sin(2k\pi/N)\right)^2}{\sum_{k=0}^{N-1} x_k^2}, \tag{9}$$

where $N$ is the length of the trajectory sample that corresponds to integration time $T = N\Delta t$. The formula (9) does not allow the simultaneous calculation for both the trajectory and the estimator $\gamma$. However, $\gamma$ should be assumed as a dynamical variable along the trajectory.

## 4 Numerical Results and Conclusions

In Fig. 3 the evolution of $\gamma(N)$, compared to the finite Liapunov characteristic number $\lambda(N)$ is shown for the same trajectories as that of Fig. 2. For the calculation of $\gamma(N)$ we used the time series $z_n = \sin(x_n)$, $n = 0, ..., N-1$ in order to avoid the discontinuity caused by the modulo operation in (2). Cases (a) and (b) clearly indicate regular dynamics. Both, $\gamma(N)$ and $\lambda(N)$ converge rapidly to zero, almost as $1/N$. In case (c), which corresponds to a weakly chaotic trajectory, we obtain convergence, for both $\gamma(N)$ and $\lambda(N)$, which holds during that time the orbit is sticky. Afterwards, $\lambda(N)$ tends slowly to saturation, while $\gamma(N)$ shows an abrupt increment indicating that the trajectory entered the chaotic channel and reveals its chaotic nature efficiently. In case (d) $\lambda(N)$ seems to tend at a relatively large value indicating the chaotic character of the trajectory. The estimator $\gamma(N)$ shows a remarkable divergence indicating, additionally, a slow diffusion that is apparent at least up to the integration interval. Independently of the parameter $k$, cases (a) and (b) are typical for all regular trajectories, and, in average, it holds $\gamma \sim N^a$ with $a \approx 1$. It worths to note, that the convergent evolution of $\gamma(N)$ is followed by dense sharp peaks, towards to lower values. Such peaks indicate that the corresponding spectra should be

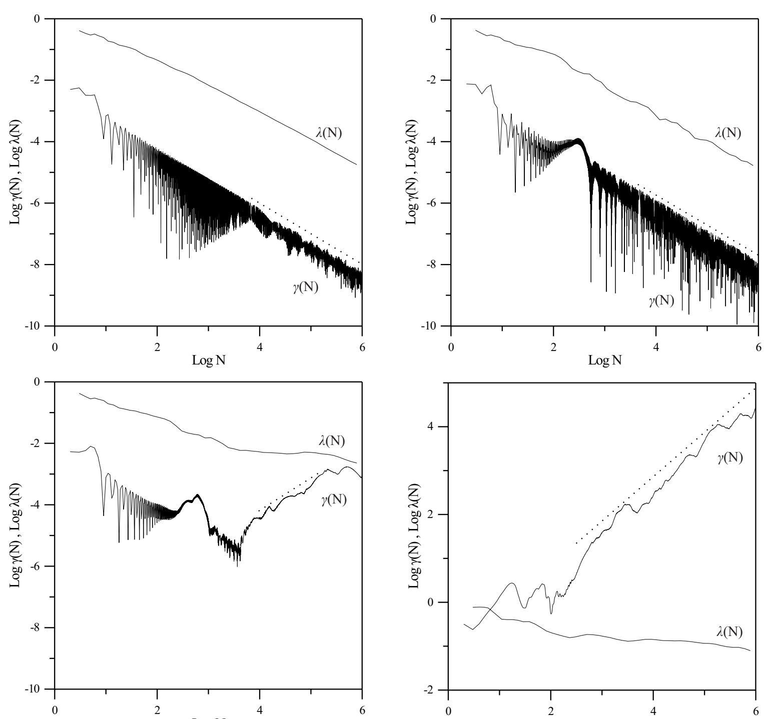



**Fig. 3.** The evolution of $\gamma(N)$ and the finite time Liapunov characteristic number for the trajectories referred in Fig. 2.

discrete as, indeed, they are. For chaotic trajectories, the evolution of $\gamma(N)$ is quite smooth because of the continuity of the associated spectra.

From Fig. 3 we observe that $\gamma(N)$ and $\lambda(N)$ show the same asymptotic characteristics that can be used to classify the trajectories as ordered or chaotic. Especially for ordered trajectories, $\gamma(N)$ and $\lambda(N)$ show a similar decay with respect to $N$. For chaotic orbits, $\gamma(N)$ and $\lambda(N)$ seem to reach saturation values in different ways, but nevertheless, we can conclude their non-convergent character. In Fig. 4 we plot the pairs $(\gamma(N), \lambda(N))$ obtained from 2000 randomly selected initial conditions for $k = 0.9$ (crosses) and $k = 1$ (circles) and after $N = 3 \cdot 10^6$ iterations. The 97% of the points belongs either in the domain A (regular trajectories) or in the domain B (chaotic trajectories). The rest of them, which mainly correspond to $k = 0.9$, may be classified as weakly chaotic trajectories. By increasing the number of iterations, the points in region A, and, generally, the points that correspond to ordered motion, move to lower values, since both $\gamma(N)$ and $\lambda(N)$ tend to zero. Therefore, the reliability of the trajectory classification increases as the length of the trajectory increases.

In many cases, $\gamma$ shows in average $N^a$-divergence, with $a \approx 1$. However, after a long time evolution, $\gamma$ shows small oscillations around a constant value $\bar{\gamma}$ (saturation). The value of $\bar{\gamma}$ is proportional to how long the $N^a$-divergence takes place and it is associated to the $1/f^a$-divergence of the spectrum caused by the slow diffusion through small islands and cantori [8]. Thus, large values of $\bar{\gamma}$ indicate the diffusive character of the trajectory rather than its strong chaotic evolution. In Fig. 5 we can see that the chaotic orbits, which start near the unstable fixed point (0,0) for $k \leq 1$, correspond to large values for $\bar{\gamma}$ (about $10^4$), but, as $k$ increases, we observe that $\bar{\gamma} \to 1$. According to the definition of $\gamma$, this value is obtained when the power spectrum has the form $p(f) \approx$const (white noise).

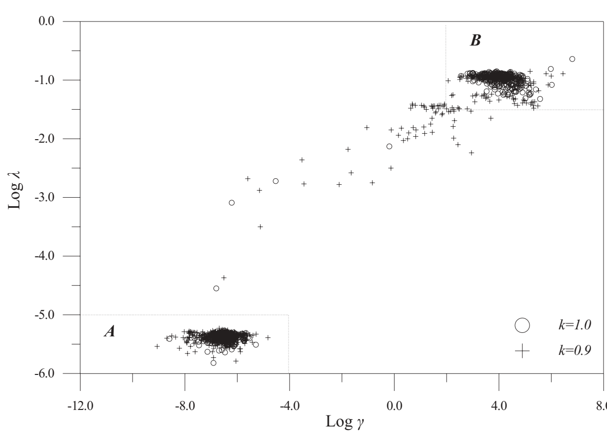



**Fig. 4.** The $(\gamma(N)$,$\lambda(N))$ graph for 2000 trajectories of the standard map. Points that belong to regions A or B indicates regular and chaotic dynamics, respectively, with great certainty.

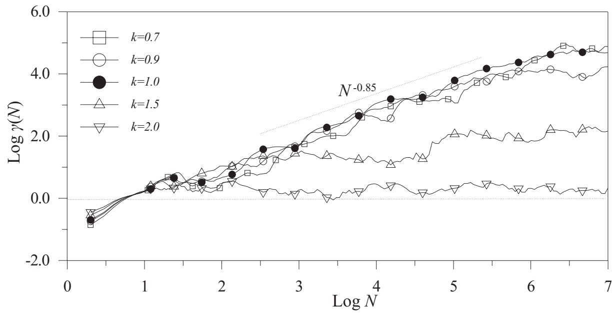



**Fig. 5.** The $\gamma(N)$ evolution along a chaotic trajectory close to the unstable fixed point (0,0) and for different values of the mapping parameter $k$.

Conclusively, we may claim that regular dynamics is associated with convergent spectra, in the sense that $p(f) \to 0$ as $f \to 0$ or, equivalently, $\gamma(T) \to 0$ as $T \to \infty$. Chaos is associated with divergent spectra where $p(f) \to$ const.$> 0$ as $f \to 0$ or, equivalently, $\gamma(T) \to$ const.$> 0$ as $T \to \infty$. In other words, for Hamiltonian bounded trajectories, discrete spectra are convergent, while, spectra with continuous local domains are divergent. Although there is not a rigorous proof, the numerical results support strongly the above relation.

The low frequency power estimator $\gamma$ can be used as an indicator for the qualitative character of a trajectory, in a similar manner to that of the Liapunov characteristic number $\lambda$. These two quantities are expressed in the same units ($\text{time}^{-1}$) but they have different physical meaning. Namely, $\gamma$ is associated with the long term regular or irregular evolution rather than the linear or exponential divergence of nearby trajectories and is calculated along a single trajectory.

**Acknowledgements**

The author would like to thank Dr. S.Ichtiaroglou for the fruitful discussions about this topic.

## References

1. V.V.Beloshapkin, G.M.Zaslavsky: Phys. Lett. A **97**, 121 (1983)
2. R.S. Dumont, P. Brumer: J. Chem. Phys. **88**, 1481 (1988)
3. T. Geisel, A.Zacherl, G.Rabons: Phys. Rev. Lett. **59**, 2503 (1987)
4. S. Kim, S. Ostlund, G.Yu: Physica D **31**, 117 (1988)

23. J. Laskar: Physica D **67**, 257 (1993)

6. S.Ostlund, D.Rand, J. Sethna, E. Siggia: Physica D **8**, 303 (1983)
7. R.Roy, B.G.Sumpter, G.a.Pfeffer, S.K. Gray, D.W. Noid: Physics Rep. **205**, 111 (1991)
8. G.E. Powell: Regular and irregular frequency spectra. Ph.D.Thesis , Queen Mary College, London (1980)
9. T.R. Press, S.A. Teukolsky, W.T. Vetterling, B.P.Flannery: *Numerical Recipes in C*, (Cambridge University Press, Cambridge, 1992)

10. R.Z. Sagdeev, D.A. Usikov, G.M. Zaslavsky: *Nonlinear Physics* (Harwood Academic 1988)
11. M. Tabor: *Chaos and integrability in nonlinear dynamics*, (Wiley, New York 1989)
12. G. Voyatzis, S. Ichtiaroglou: J. Phys. A **25**, 5931 (1992)

# Disk-Crossing Orbits

Chris Hunter

Department of Mathematics, Florida State University, Tallahassee FL 32306-4510, USA

**Abstract.** We study orbits in simplified models of galaxies which consist of two components, a disk and a halo. The disk is idealized as razor-thin, though we present evidence that this simplifying assumption is not critical. We find that the presence of the disk causes many more resonances than have been found in similar smooth potentials. Those resonances grow at relatively modest values of energy, overlap, and give rise to many stochastic orbits. A significant range of regular orbits remain and show smooth KAM curves, even though the discontinuous potential due to the razor-thin disk means that current versions of the KAM theorem do not apply.

## 1 Introduction

Many galaxies have a disk component and one or more other components which are much less flattened. Normal spirals and S0s are examples. Bender et al [1] have found that many ellipticals can be classified as disky, and that their diskiness is consistent with a disk plus bulge model [2]. For simplicity, we use halo as an all-embracing term to describe the whole non-disk component. Orbits of halo stars in such galaxies will necessarily cross back and forth through the disk. As they do, they will experience a fairly abrupt change in the gravitational force field. This paper examines how these changes in the force field affect the dynamics of the orbit. It is a topic which seems to have attracted very little attention so far, with two notable exceptions. Ostriker, Spitzer, and Chevalier [3] have discussed the effect that the compressive gravitational shocks, caused by passage through the disk, have on the internal structure of globular clusters. Our interest is on how that passage affects the orbits of individual stars of the halo population, on the resonances which they can cause, and on the extent to which they can induce chaos. This issue was also studied in the 1970s by L. Martinet and co-authors [4]–[7] who integrated orbits in Schmidt's [8], [9] two models of the Galaxy. These models have highly flattened disk components.

We shall use simple models which have both disk and halo components. These are the Kuzmin-like potentials which were introduced recently by Tohline and Voyages [10]. Their disk components are razor-thin. We describe their potentials and densities in Sect. 2, and their dynamics in Sect. 3. Numerical integrations are needed to study the detailed properties of orbits. We describe the results of those integrations in Sect. 4. In Sect. 5 we show, again by means of orbit integrations, that our findings do not depend critically on the simplifying approximation that the disk component is razor-thin. Sect. 6 sums up, and relates our findings to other work.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 137–153, 2003.

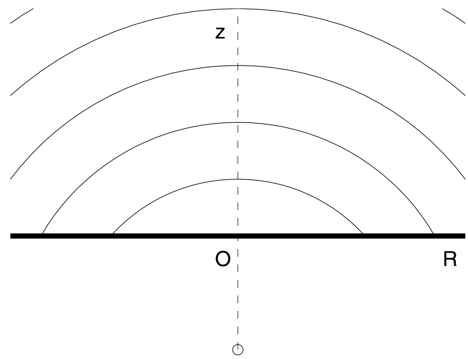



**Fig. 1.** The thin solid lines in $z > 0$ are equipotential surfaces above Kuzmin's disk. They are circles in $(R, z)$-space centered on an image mass at $z = -a$, which is shown as a small open circle. The equipotentials in $z < 0$ are their reflections in the disk $z = 0$

## 2 Models

### 2.1 Kuzmin's Disk

Kuzmin's disk [11], [3] is a flat density distribution with the remarkable property that the gravitational potential above it is that due to a point mass at a distance $a$ below it, while the potential below it is that due to a point mass at a distance $a$ above it. Its gravitational potential is

$$\Phi = \frac{-GM}{\sqrt{R^2 + (a + |z|)^2}}, \tag{1}$$

when $z = 0$ is the plane of the disk, and $R$ measures radial distance from the $z$-axis of symmetry. Its equipotential surfaces are spherical, those in $z > 0$ centered on the point mass below the disk, and those in $z < 0$ on the point mass above the disk.

### 2.2 Kuzmin-Like Potentials

Kuzmin's disk is a particular case of the more general Kuzmin-like potentials of Tohline and Voyages [10], for which

$$\Phi = \Phi(\xi), \qquad \xi = \sqrt{R^2 + (a + |z|)^2}. \tag{2}$$

They have the same spherical equipotential surfaces as Kuzmin's disk, with centers of force on the opposite side of the $z = 0$ plane, but a force law which is more general than the inverse-square of Kuzmin's disk. The variable $\xi$ measures distance from the centers of force. Poisson's equation shows that the potential (2) is generated by a density

$$\varrho = \frac{1}{4\pi G}\Delta\Phi = \frac{1}{4\pi G}\left[\Phi''(\xi) + \frac{2[1 + a\delta(z)]}{\xi}\Phi'(\xi)\right]. \tag{3}$$

The Dirac delta function $\delta(z)$ arises from derivatives of $|z|$ because $\partial|z|/\partial z = \mathrm{sgn}(z)$, and $\partial\,\mathrm{sgn}(z)/\partial z = 2\delta(z)$. Kuzmin-like potentials therefore arise from a volume density which is stratified on the equipotentials, together with a surface density $\sigma = a\Phi'(\xi)/2\pi G\xi$ on the plane $z = 0$. It is this surface density which causes the $z$-component of force to be discontinuous. The volume density vanishes for the special case of Kuzmin's disk, and there is then only a surface density $\varrho = aM\delta(z)/2\pi(a^2 + R^2)^{3/2}$.

The choice of a specific spherical potential $\Phi(\xi)$ fixes both the volume and the surface densities. One way of judging their relative significance is to compute the mass of each component interior to some equipotential $\xi = \xi_0$. Using the geometric formula of $4\pi\xi(\xi - a)$ for the surface area of the two part-spheres which form the equipotential of constant $\xi$, we obtain

$$M_{\mathrm{disk}}(\xi_0) = \frac{a}{G}\left[\Phi(\xi_0) - \Phi(a)\right], \qquad M_{\mathrm{total}}(\xi_0) = \frac{\xi_0}{G}(\xi_0 - a)\Phi'(\xi_0), \tag{4}$$

for the disk and total mass interior to the equipotential $\xi = \xi_0$.

### 2.3 The Logarithmic Kuzmin-Like Potentials

As an example, we shall investigate the logarithmic potential which is widely used as a model in galactic dynamics [3]. It gives a Kuzmin-like model with potential and density

$$\Phi(\xi) = V_0^2 \ln \xi, \quad \varrho = \frac{V_0^2}{4\pi G\xi^2} + \frac{aV_0^2\delta(z)}{2\pi G(R^2 + a^2)}. \tag{5}$$

As (4) shows, the mass in the disk grows logarithmically with increasing distance whereas the total mass grows linearly. The ratio of these masses is

$$\frac{M_{\mathrm{disk}}(\xi_0)}{M_{\mathrm{total}}(\xi_0)} = \frac{a}{(\xi_0 - a)} \ln\left(\frac{\xi_0}{a}\right), \tag{6}$$

and the relative significance of the disk diminishes with increasing $\xi_0$. Orbits at higher energies and greater distances experience forces for which the halo is increasingly dominant. Figure 2-a shows how isophotes would appear when viewed at a small angle from edge-on, assuming the same mass-to-light ratio for all mass.

## 3 Dynamics

Motion in a spherical potential conserves the energy $E$ and the angular momentum vector about the center. The motion is integrable and is confined to the plane through the center of force that is perpendicular to the angular momentum vector. Unless confined always to the plane $z = 0$, an orbit in a Kuzmin-like potential passes continually back and forth between the regions $z > 0$ and $z < 0$,

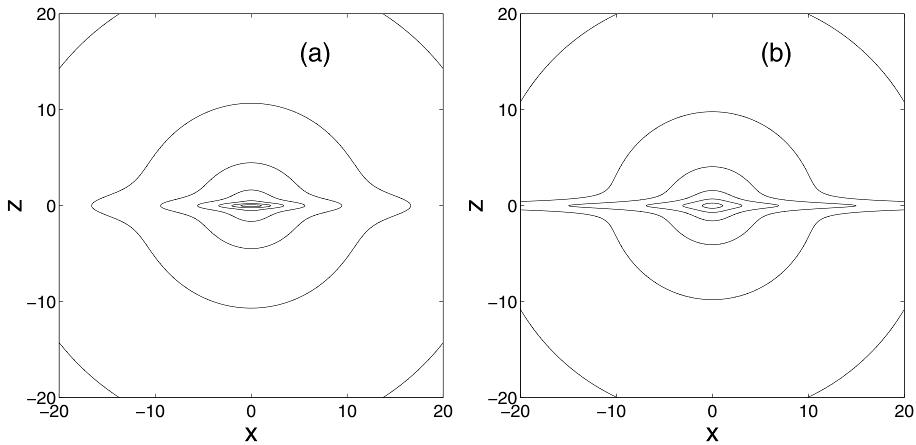



**Fig. 2.** (a) Projected density of a logarithmic Kuzmin-like model (5), assuming the same mass-to-light ratio for all mass, when viewed at 5° from edge-on (left). Contour levels decrease by factors of $10^{-1/3}$. The projected disk density is infinite when viewed edge-on, and the isophotes become increasingly less pointy as the viewing angle increases. (b) Projected density of the same model after thickening with $b = 0.5$ as described by (12) in Sect. 5 when viewed edge-on. The centermost contours and successive outer ones represent the same levels as in (a). Thickening spreads density out from the disk, giving higher volume densities; the level of the outermost contour of (b) matches the next to outermost of (a)

and hence from one spherical potential field to the other. Different angular momenta are conserved in the two regions. Both are described by the vector

$$\mathbf{J} = [\mathbf{r} + a\,\mathrm{sgn}(z)\mathbf{k}] \times \mathbf{v}, \tag{7}$$

where $\mathbf{r}$ is the position relative to the origin, $\mathbf{v}$ is velocity, and $\mathbf{k}$ is the unit vector in the $z$-direction. An equation of motion for $\mathbf{J}$ can be derived from the equations of motion in the Kuzmin-like potential. They are

$$\frac{\mathrm{d}\mathbf{r}}{\mathrm{d}t} = \mathbf{v}, \qquad \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} = -\nabla\Phi = -\frac{[\mathbf{r} + a\,\mathrm{sgn}(z)\mathbf{k}]}{\xi}\Phi'(\xi). \tag{8}$$

It follows that

$$\frac{\mathrm{d}\mathbf{J}}{\mathrm{d}t} = [\mathbf{r} + a\,\mathrm{sgn}(z)\mathbf{k}] \times \frac{\mathrm{d}\mathbf{v}}{\mathrm{d}t} + \mathbf{v}\times\mathbf{v} + 2a\delta(z)(\mathbf{k}\times\mathbf{v})\frac{\mathrm{d}z}{\mathrm{d}t} = 2a\delta(z)(\mathbf{k}\times\mathbf{v})\frac{\mathrm{d}z}{\mathrm{d}t}. \tag{9}$$

This equation confirms the constancy of the vector $\mathbf{J}$ except when the orbit crosses the disk. Then $\mathbf{J}$ changes discontinuously by an amount $\pm 2a(\mathbf{k}\times\mathbf{v})$ unless $\mathbf{k}\times\mathbf{v} = 0$, that is unless the orbit is then travelling perpendicularly to the disk. The change in $\mathbf{J}$ is always perpendicular to the $z$-direction because the $z$-component $J_z$ of $\mathbf{J}$ is conserved; $J_z$ is the angular momentum about the axis of symmetry, and is a constant of the motion, as in any axisymmetric potential.

Since it is known that energy and angular momenta are the only constants of motion in general spherical potentials and, as we have seen, neither $J_x$ nor $J_y$ are conserved for general Kuzmin-like potentials, it follows that, contrary to what is stated in [10], Kuzmin-like potentials are not generally integrable. Kuzmin's disk is exceptional, its potential is Stäckel, and so all orbits in it are integrable. Their integrability relies on the Keplerian nature of its spherical potentials; motions then have an extra integral of motion, the Laplace-Runge-Lenz vector [13], in addition to energy and angular momentum. The Laplace-Runge-Lenz vector changes discontinuously across the disk too. However, there is a linear combination of its $z$-component and $\mathbf{J}^2$ for which the two discontinuities cancel [14], and this combination provides Kuzmin's third integral [11]. (Another exceptional case is that of a harmonic potential $\Phi \propto \xi^2$. It is not generally relevant for galaxies, except as an approximation for orbits which remain close to the center of the disk.)

Because position and velocity of an orbit are continuous as the disk is crossed, the tangent to an orbit changes smoothly. The hallmark of the discontinuity in the gravitational force is a discontinuity in the curvature of an orbit. Fig. 3-a gives the clearest example of this feature. It shows a particular $J_z = 0$ periodic polar orbit in the potential due to Kuzmin's disk. The orbit in $z < 0$ is part of an ellipse with focus at the center of force in $z > 0$. This ellipse is oriented such that the orbit is travelling directly away from the center of force in $z < 0$ as it crosses to $z > 0$. Hence its trajectory in $z > 0$ remains a straight line away from the lower center of force. The orbit travels along that line until it reaches the maximum extent which its energy allows, after which it returns and retraces its path.

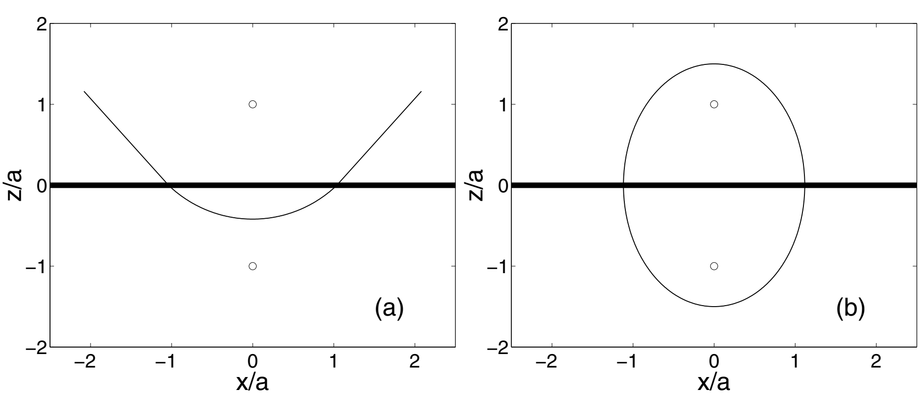



**Fig. 3.** The simplest periodic disk-crossing polar orbits for Kuzmin's disk: (a) a box orbit, which is a banana in the terminology of Binney [15] and a boxlet in the terminology of Miralda-Escudé and Schwarzschild [16], and (b) an ellipse, for energy $E = -GM/3a$. The open circles again denote the image masses as in Fig. 1. Similarly shaped periodic orbits occur for other Kuzmin-like potentials. The banana boxlet has a twin, which is its mirror image in $z = 0$.

# 4 Orbits

## 4.1 Orbits in Kuzmin's Disk

Orbits with $J_z \neq 0$ are short-axis tubes. Polar orbits with $J_z = 0$, which lie in planes through the axis of symmetry, may be either boxes or loops. At low energies $-1 < E < -GM/2a$ for which the orbit along the $z$-axis is stable, only box orbits occur. Loop orbits become possible at higher energies in the range $-GM/2a < E < 0$ for which the $z$-axis orbit has become unstable [14] The parent of the loop orbit family is a closed ellipse, as shown in Fig. 3-b. It is not quite a Keplerian ellipse because the orbit is always under the influence of a point mass at the more distant focus. Rather it is the union of two slower halves of a Keplerian ellipse. Although there are box orbits for all energies, the type of banana boxlet shown in Fig. 3-a occurs only for the energy range $-GM/2\sqrt{2}a \leq E < 0$. The curved part of the orbit is a quarter circle for the lower limit of energy, and so is nearly circular for the case illustrated. Its eccentricity increases as $E$ increases, and the linear parts, which extend to a distance $-GM/E$ from the lower center of force, lengthen.

## 4.2 Orbits in Kuzmin-Like Potentials

In the following subsections, we present numerical results for the logarithmic potential (5) using $z = 0$ surfaces of section. Surfaces of section for Kuzmin's disk can be found analytically as contours on which the third integral is constant, but numerical integration is needed for Kuzmin-like potentials. We computed orbits by a variant of the method described in Appendix A.1 of Hunter et al [17] which integrates accurately from one crossing of $z = 0$ to the next using polar coordinates in $(z, v_z)$-space. No interpolation is used to find when crossings occur, and there is no uncertainty as to which force formula to use. The surfaces of section for the logarithmic potential (5) contain resonant island chains and stochastic as well as regular regions. They stand in marked contrast to the regular surfaces of section for Kuzmin's disk, for which the only significant feature is the bifurcation when loops first appear with polar orbits. Surfaces of section are plotted for specific values of energy and angular momentum $J_z$ about the axis of symmetry. We express $J_z$ as a fraction $k$ of the angular momentum of the circular orbit in the $z$-plane., i.e. the maximum possible $J_z$ for the given energy. We use $a$ as unit of length, and for ease of understanding, identify energies by $R_c$, the radius of a circular orbit at that energy. In terms of $R_c$, the energy is

$$E = \frac{1}{2}\left[\ln(R_c^2 + 1) + \frac{R_c^2}{R_c^2 + 1}\right]. \tag{10}$$

For the unbounded potential (5), $E \geq 0$, and $R_c$ increases monotonically with the energy, first as $R_c \approx \sqrt{E}$ for small $E$, but later as $R_c \approx e^E$ for large $E$. We label periodic orbits by the ratio $m : n$ where $m$ and $n$ are the numbers of the full cycles in $z$ and $R$ (or $x$ in the case of polar orbits) respectively, in a complete cycle.

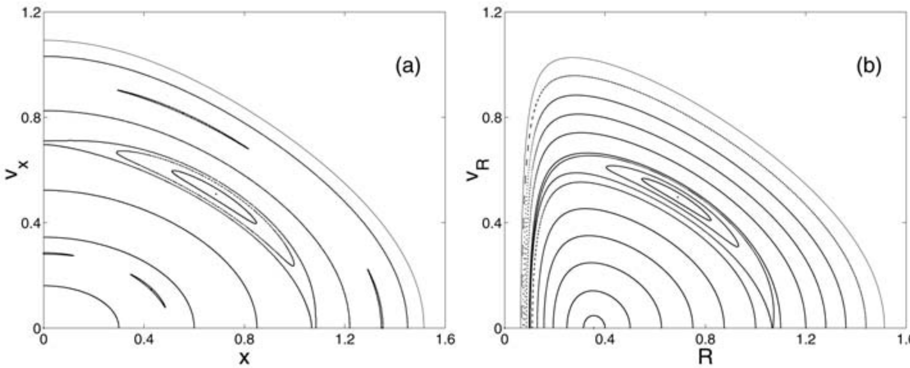



**Fig. 4.** (a) Surface of section for polar orbits in the logarithmic potential (5) for $R_c = 1$, (b) Surface of section for axisymmetric orbits in the same potential and at the same energy, and with $k = 0.1$, i.e. angular momentum 0.1 of that for a circular orbit at this energy

## Polar Orbits

Only box orbits occur at low energies, and $(x, v_x)$ surfaces of section mostly show smooth rings centered on the $z$-axis orbit at $(0, 0)$. However bifurcations do occur before the $z$-axis orbit becomes unstable at $R_c = 1.37$. Figure 4-a for $R_c = 1$ shows three rings of island chains. The prominent middle ring is due to the $2:1$ banana boxlets, and their unstable figure-of-eight companion. These orbits bifurcate from the $z$-axis orbit when $R_c = 0.62$. Initially their paths lie close to the $z$-axis, but their initial island chain cuts the $z$-axis at $z = 0.35$, and is clearly removed from $(0, 0)$. The smooth curves surrounding $(0, 0)$ remain, indicating that the stability of the $z$-axis orbit is unaffected by this bifurcation. The outer narrow ring corresponds to $3:1$ orbits. They bifurcate from the $z$-axis orbit at a lower energy when $R_c < 0.45$. The inner narrow ring corresponds to $3:2$ orbits of a well-known type; an unstable fish and a stable antifish [16]. A separatrix between box and loop orbits is formed once the $z$-axis orbit becomes unstable at $R_c = 1.37$, and a visible stochastic region develops around the separatrix [18]. Figure 5 shows that resonant islands take up a significant part of the box orbit region at $R_c = 3$. The $2:1$ banana boxlet at $x = 1.84$, $v_x = 1.15$ is still prominent, while the $3:2$ family with periodic points at $x = 2.54$, $v_x = 0.79$ and $x = 0$, $v_x = 1.14$ has grown in significance. Figure 6 shows that most of the box orbit region has become stochastic when energy has increased to $R_c = 10$, as a result of substantial overlapping of the resonant regions which are so prominent in Fig. 5. There are many resonant islands in Fig. 6, and even islands within islands, but the stochastic sea dominates the outer part of the surface of section. The island for the banana boxlets, which is here centered on $x = 3.2$, $v_x = 1.7$, has moved further outwards and become thinner. The outward migration is due to the fact that these orbits become flatter with increasing energy; the outer boundary of the surfaces of section is the orbit which is always in the plane

$z = 0$. A few box orbits remain in the outer parts of the surface of section. They too are flat and never depart far from the $z = 0$ plane.

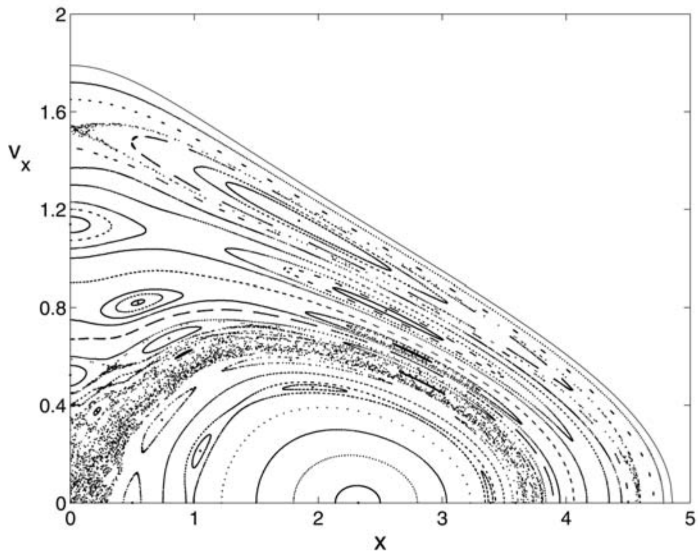



**Fig. 5.** Surface of section for polar orbits in the logarithmic potential (5) for $R_c = 3$

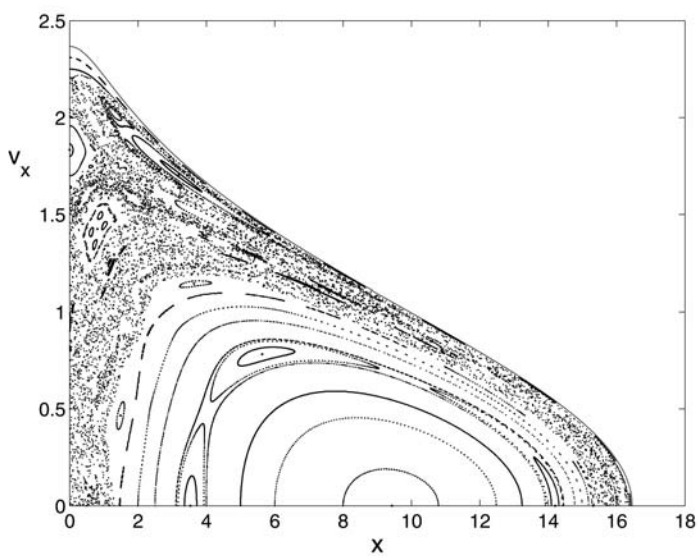



**Fig. 6.** Surface of section for polar orbits in the logarithmic potential (5) for $R_c = 10$

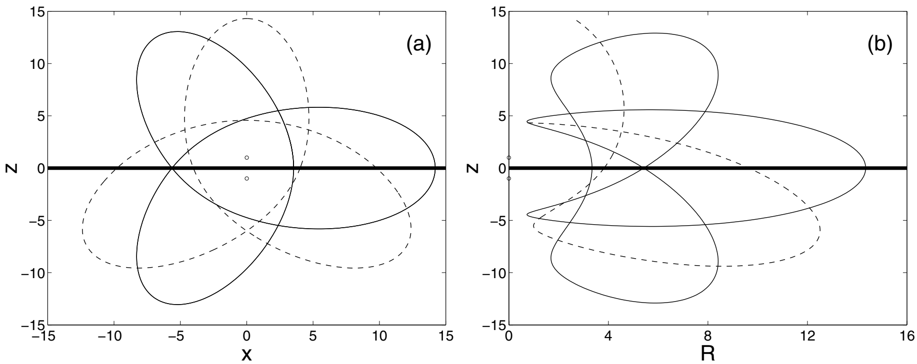



**Fig. 7.** (a) Stable and unstable 2 : 2 orbits, full and broken lines respectively, which occupy the centermost chain of islands in Fig. 6. The reflection of the unstable orbit in $z = 0$ gives another unstable 2 : 2 orbit. (b) The stable and unstable 2 : 4 orbits, full and broken lines respectively, which occupy the centermost chain of islands in Fig. 11

While the box orbit region is becoming stochastic, the loop orbit region, centered around the 1 : 1 periodic closed loop, remains largely regular, though with some resonant islands. The first out from the center and most prominent in Fig. 6 are due to the 2 : 2 periodic orbits whose three-petalled rosette form is shown in Fig. 7-a.

**Delay Plots of Angular Momentum** provide a third way of analyzing orbits in Kuzmin-like potentials, which supplement plots in position space and surfaces of section. While the orbit is above the disk, $J_y$ is its angular momentum about the lower center of force, and remains constant. The next time the orbit returns to $z > 0$, it has a new value of $J_y$, except for exceptional cases such as the elliptical orbit of Fig. 3-b. The sequence of successive values of $J_y$ characterizes an orbit, and generates a delay plot, similar to those used for analyzing time series generated by other dynamical systems [19] [20], in which two successive values are plotted as a point, the earlier value giving the abscissa and its successor the ordinate. Figure 8 displays $J_y$ delay plots for four orbits which are included in the surface of section of Fig. 6. The physical forms of the first three are shown in Fig. 9. The orbits are arranged by decreasing stochasticity and increasing regularity. Orbit (a) inhabits the stochastic region just outside the regular central core of loop orbits. The points in its delay plot are scattered fairly uniformly. Figure 9-a shows that it almost fills the area within the zero-velocity curve, apart from narrow slivers near the left and right edges. Orbit (b) from well inside the stochastic region fills less than half the region within the zero-velocity curve. Its delay plot shows much randomness, but also a double ring structure. Orbit (c) is something of a surprise because it seems to start in a stochastic region, yet its delay plot is highly ordered. Its pretzel shape as seen in Fig. 9-c shows it to belong to the family of 4 : 3 periodic orbits parented by the periodic orbit which crosses the surface of section of Fig. 6 at $(0.81, 1.42)$ and $(6.38, 1.21)$. Orbit (c)

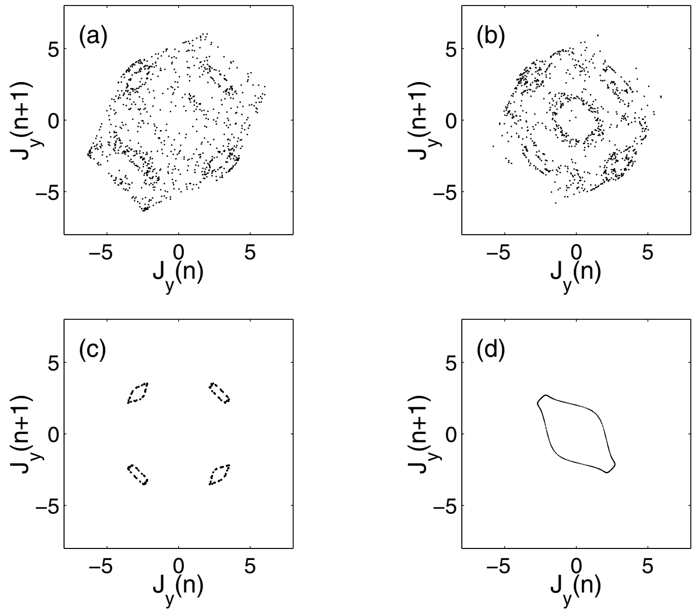



**Fig. 8.** $J_y$ delay plots for four $R_c = 10$ polar orbits. $J_y(n)$ is the value of $J_y$ during the $n$-th passage through $z > 0$. The orbits start in $z = 0$ from (a) $x = 2$, $v_x = 1$, (b) $x = 1$, $v_x = 1.7$, (c) $x = 4$, $v_x = 1.4$, (d) $x = 0$, $v_x = 2.25$

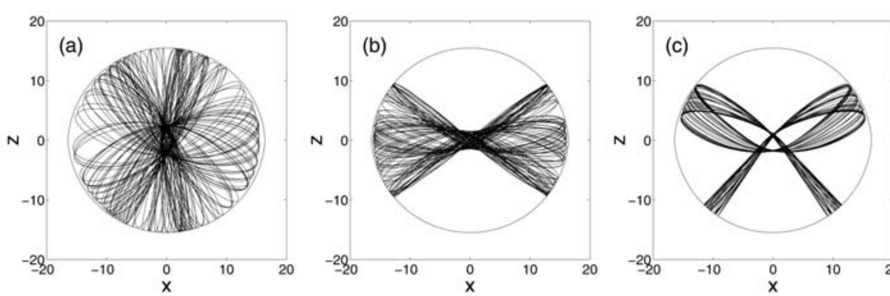



**Fig. 9.** Orbits for first three delay plots of Fig. 8, after 200 crossings of $z = 0$. The outer boundary is the zero-velocity curve

generates the set of broken lines which surrounds the four island chain around $(0.81, 1.42)$, and broken lines appear in its delay plot too. The starting point of this orbit at $(4, 1.4)$ lies at the left end of a long and narrow island surrounding the other periodic point at $(6.38, 1.21)$. That island is lost in the complexity of Fig. 6. Finally, orbit (d) is a regular box orbit which gives a smooth curve around the outer part of the surface of section. Its delay plot is also smooth and regular. Being close to the outer boundary, this orbit is a flat box which stays close to the disk at all times.

**Axisymmetric Orbits**

The surfaces of section Fig. 4-b, Fig. 10, and Fig. 11 for the same three energies of $R_c = 1$, 3, and 10, respectively, and for $J_z$ angular momenta which are one-

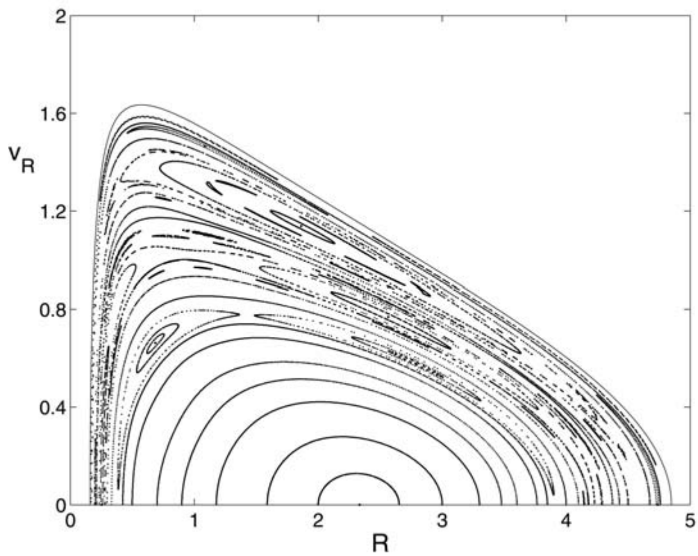



**Fig. 10.** Surface of section for axisymmetric orbits in the logarithmic potential (5) for $R_c = 3$ and $k = 0.1$

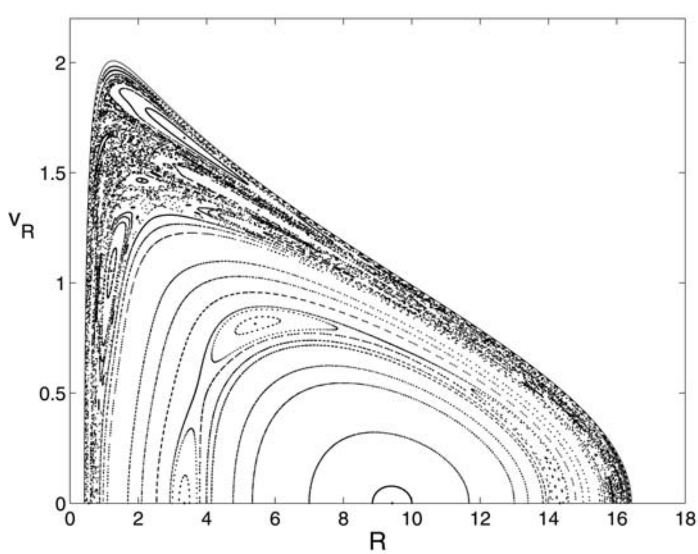



**Fig. 11.** Surface of section for axisymmetric orbits in the logarithmic potential (5) for $R_c = 10$ and $k = 0.1$

tenth of the maximum possible ($k = 0.1$) show similar trends with increasing energy to the corresponding figures for polar orbits. The predominant class of regular orbits are short axis tubes [14]. They give the smooth rings surrounding the center which represents a thin tube or shell. Figure 4-b shows a prominent island at the lowest energy. This is due to the 1 : 1 saucer orbit [21] which is obtained when the angular momentum barrier stops a banana orbit before it gets to the $z$-axis and reflects it back along its path after its first crossing of the disk. The $R_c = 3$ surface of section in Fig. 10 shows its outer part to be occupied by resonant island chains. Those resonances have overlapped when the energy has increased to $R_c = 10$, and most of the outer part is stochastic. However there is a large regular core with a prominent island chain. That chain corresponds to the axisymmetric counterparts of the 2 : 2 rosettes of the polar case. The angular momentum barrier truncates and reflects these orbits to give the 2 : 4 orbits shown in Fig. 7-b, a stable spaceship and an unstable pair of twisted fish. Note that this island chain is present already as the innermost island chain in Fig. 10 for the lower energy of $R_c = 3$. There is a difference in that there the unsymmetric twisted fish are stable, and the spaceship is unstable.

Delay plots can be constructed for this case too, using the quantity $J_x^2 + J_y^2$ which is now constant during each passage through $z > 0$. Figure 12 shows delay plots for two orbits from the outer stochastic part of Fig. 11. The left delay plot shows wide scattering, with some concentration at the boundaries, and an empty hole at the top right, while the right plot shows far less scattering. Figure 13 shows that neither orbit comes close to filling the region within the zero-velocity curve. Orbit (b) with the more organized delay plot, ranges more widely, but less randomly, than orbit (a).

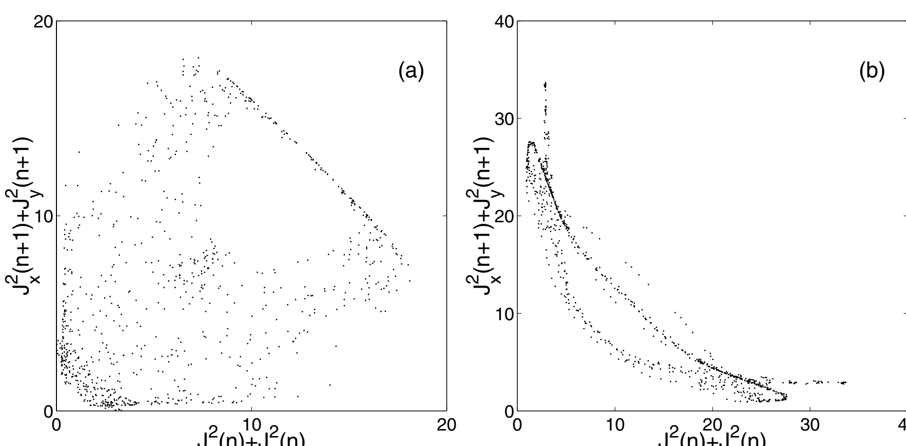



**Fig. 12.** $J_x^2 + J_y^2$ delay plots for two $R_c = 10$, $k = 0.1$ axisymmetric orbits. The orbits start in $z = 0$ from (a) $R = 1$, $v_R = 1.7$, (b) $R = 1.6$, $v_R = 1

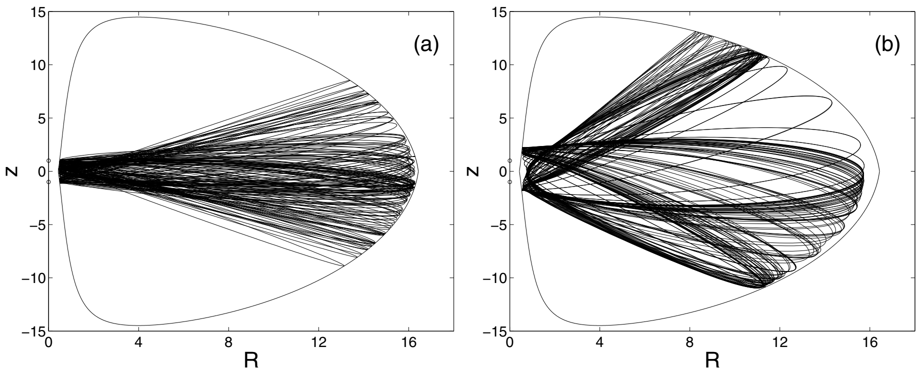



**Fig. 13.** Orbits for the two delay plots of Fig. 12 after 200 crossings of $z = 0$. The outer boundary is the zero-velocity curve

## 5 Thickened Disks

We have so far been investigating force fields which have sharp discontinuities. That is an idealization of a galaxy with a disk-like component, and so one must wonder how critically do the results which we have found depend on the assumption of an abrupt discontinuity. We investigate this issue by thickening the disk, borrowing a device introduced by Miyamoto and Nagai [22]. We add a length $b$, which is zero in the limit of a razor-thin disk, to the definition of the variable $\xi$ on which the potential depends, so that now

$$\Phi = \Phi(\xi), \qquad \xi = \sqrt{R^2 + (a + \sqrt{z^2 + b^2})^2}. \tag{11}$$

This potential is due to the density field

$$\varrho = \frac{1}{4\pi G}\left[\Phi'' + \frac{2\Phi'}{\xi} + \frac{ab^2\Phi'}{\xi(z^2+b^2)^{3/2}} - \frac{b^2}{\xi}\left(\frac{\Phi''}{\xi} - \frac{\Phi'}{\xi^2}\right)\left(1 + \frac{a}{\sqrt{z^2+b^2}}\right)^2\right] \tag{12}$$

The first three components give the density (3) in the limit $b \to 0$ when $b^2/(z^2 + b^2)^{3/2} \to 2\delta(z)$. The right half of Fig. 2 shows contours of the projected density (11) for the logarithmic potential of (5) when viewed edge-on. Comparison with the unthickened case in the left half shows that there has been a considerable lowering of density contrasts. Even though $\xi$ can no longer be interpreted as a specific distance, it does tend to a radial distance at large distances where most of the equipotentials become quite spherical. Despite the changes in the potential for the displayed case of $b = 0.5$, which is not much smaller than the displacements $a = 1$ of the former centers of force, yet the $R_c = 10$ surface of section shown in Fig. 14 shows no significant qualitative differences from the $b = 0$ case of Fig. 11. In fact much of the microstructure persists. For example, the small island around $(2.21, 1.49)$ in Fig. 14 and that around $(2.11, 1.46)$ in Fig. 11 are due to similarly shaped $7:10$ orbits, and the narrow tadpoles around $(1.11, 1.71)$ in Fig. 14 and $(0.92, 1.62)$ in Fig. 11 are due to similar $5:6$ orbits.

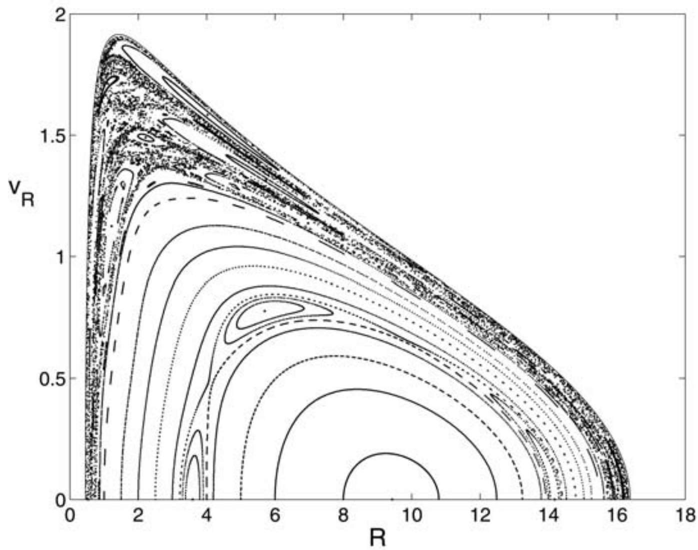



**Fig. 14.** Surface of section for axisymmetric orbits for $R_c = 10$ and $k = 0.1$, as in Fig. 11 but for a thickened potential with $b = 0.5$

# 6 Discussion and Conclusions

## 6.1 Related Work

This work adds another example to those already known in which chaos is induced when orbits pass between two or more regions, in each of which the motion is integrable, but in which different integrals apply. One that has attracted considerable attention in recent years is that of a black hole at the center of an otherwise smooth-cored triaxial galaxy. Shortly after de Zeeuw [14] had shown how Stäckel potentials account nicely for the major orbit families which Schwarzschild [23] had found in his numerical integrations of a smooth-cored triaxial galaxy, Gerhard and Binney [24] showed how destructive a central black hole, or cusp, would be for box orbits which came close to it. Motion near the black hole is nearly Keplerian, as has been discussed in detail by Sridhar and Touma [25].

Transitions between regions in which different integrals apply are basic to the mapping constructed by Wisdom [26] to explain the Kirkwood gaps in the asteroid belt. His mapping concentrates high-frequency perturbations by Jupiter into a periodic sequence of impulses which modify the motion periodically. Between impulses, the motion is determined entirely by secular terms which describe an integrable system. Touma and Tremaine [27] have applied a similar mapping in galactic dynamics. They study eccentric orbits in a plane non-axisymmetric potential using a map obtained by concentrating the non-axisymmetry so that it acts impulsively at apocenters of orbits. Between successive apocenters the non-axisymmetry is ignored and the orbital motion is integrable.

From a mathematical perspective, this work adds another instance in which there is evidence that KAM curves occur, even though the conditions for which the KAM theorem has been proved are not satisfied. Moser's [28] proof of that theorem has been refined to the case of continuous derivatives through to third order. Yet Kuzmin-like potentials have discontinuous first derivatives. Like the example of Varvoglis [29], ours comes directly from a significant physical context. Many studies in which KAM curves have been found for Hamiltonians which lack the smoothness required by Moser's proof, such as that of Benettin and Strelcyn [30] have studied variants of the billiard ball problem introduced by Birkhoff [31].

### 6.2 Conclusions

Our largely numerical study of orbits in Kuzmin-like potentials has shown that Kuzmin's disk is not typical. Its out-of-plane orbits are all integrable, despite the abrupt changes in the angular momentum vector which occur whenever the disk is crossed. Other Kuzmin-like models (other than the simple case in which the potential is harmonic) are not integrable. The presence of the disk causes many more resonances than have been found in similar smooth potentials [32] [17]. Those resonances grow at relatively modest values of energy, overlap, and give rise to many stochastic orbits for which the disk crossings may scatter the angular momentum vector fairly randomly. The extent of the stochasticity grows with increasing energy and decreasing angular momentum. However, there are many orbits which remain regular despite the impulsive changes due to crossing the disk. We have investigated only a very limited class of Kuzmin-like models, and Kuzmin-like models are themselves special because they do not permit the disk density to be adjusted independently of the volume density. Yet it is hard to see why similar mechanisms will not induce stochasticity in disk-crossing orbits in galaxies with disks of substantial size. The similar results from our even more limited tests with thickened Kuzmin-like potentials, for which the disk is not a sharp discontinuity and for which the motion is not integrable outside the disk, is evidence of the likely robustness of our findings. Martinet et al [4]–[7] investigated quite different models. Some of their surfaces of section show large stochastic regions, more extensive than any we display, and even instances in which the central periodic orbit of an axisymmetric case has become unstable. Because the models are so different, no detailed comparison of our results with theirs is possible.

Although we drew a distinction in Sect. 1 between our interest in orbits and that of Ostriker et al [3] on the internal structure of globular clusters, that distinction may be somewhat fuzzy. The occurrence of chaos is commonly ascribed to sensitivity to initial conditions, when small changes in initial conditions lead to greatly different outcomes [20]. The calculations in [3] are based on changes in the relative motion of two nearby points of a cluster caused by disk shocking. Orbits of nearby points in a stable cluster are nearby when viewed on the galactic scale. They grow far apart if and when the cluster disrupts. Our work finds that there are substantial regions of phase space in which orbits remain regular. In them nearby orbits diverge only gradually from each other, whereas they diverge rapidly from each other in stochastic regions. This implies that the vulnerability of a globular cluster to disk shocking is likely to depend considerably on its orbit This matches what Aguilar, Hut, and Ostriker [33] indeed find; that disk shocking is much more destructive of globular clusters on highly elongated radial orbits.

**Acknowledgements**

It is a pleasure to thank George Contopoulos, and Nikos Voglis and his committee, for organizing a stimulating workshop, and for their hospitality. This work has been supported in part by the National Science Foundation through grant DMS-0104751.

## References

1. R. Bender, S. Döbereiner, C. Möllenhoff, A&AS, **74**, 385 (1988)
2. C. Scorza, R. Bender, A&A, **293**, 20 (1995)
3. J.P. Ostriker, L. Spitzer, R.A. Chevalier, ApJ, **176**, L51 (1972)
4. L. Martinet, A. Hayli, A&A, **14**, 103 (1971)
5. F. Mayer, L. Martinet, A&A, **27**, 199 (1973)
6. L. Martinet, A&A, **32**, 329 (1974)
7. L. Martinet, F. Mayer, A&A, **44**, 45 (1975)
8. M. Schmidt, Bull. Astron. Inst. Neth., **13**, 15 (1956)
9. M. Schmidt: 'Rotation parameters and distribution of mass in the Galaxy'. In: *Stars and Stellar Systems, Vol. 5* ed. by A. Blaauw, M. Schmidt (University of Chicago Press, Chicago, 1965) pp. 513-530
10. J.E. Tohline, K. Voyages, ApJ, **555**, 524 (2001)
11. G.G. Kuzmin, Astron.Zh., **33**, 27 (1956)
3. J. Binney, S. Tremaine: *Galactic Dynamics* (Princeton University Press, Princeton 1987)
13. H. Goldstein: *Classical Mechanics*, 2nd edn. (Addison-Wesley, Reading MA 1980)
14. P.T. de Zeeuw, MNRAS, **216**, 273 (1985)
15. J. Binney, MNRAS, **201**, 1 (1982)
16. J. Miralda-Escudé, M. Schwarzschild, ApJ, **339**, 752 (1989)
17. C. Hunter, B. Terzić, A.M. Burns, D. Porchia, C. Zink, Ann.N.Y. Acad. Sci., **867**, 61 (1998)
18. O.E. Gerhard, A&A, **151**, 279 (1985)
19. N.H. Packard, J.P. Crutchfield, J.D. Farmer, R.S. Shaw, Phys. Rev. Lett., **45**, 712 (1980)
20. K.T. Alligood, T.D. Sauer, J.A. Yorke: *Chaos: An Introduction to Dynamical Systems* (Springer, New York 1996)
21. M. Schwarzschild, ApJ, **409**, 563 (1993)
22. M. Miyamoto, R. Nagai, PASJ, **27**, 533 (1975)
23. M. Schwarzschild, ApJ, **232**, 236 (1979)
24. O.E. Gerhard, J.J. Binney, MNRAS, **216**, 467 (1985)
25. S. Sridhar, J. Touma, MNRAS, **303**, 483 (1999)
26. J. Wisdom, AJ, **87**, 577 (1982)

27. J. Touma, S. Tremaine, MNRAS, **292**, 905 (1997)
28. J. Moser: *Stable and Random Motions in Dynamical Systems* (Princeton University Press, Princeton 2001)
29. H. Varvoglis, J. Physique, **46**, 495 (1985)
30. G. Benettin, J.-M. Strelcyn, Phys. Rev. A **17**, 773 (1978)
31. G.D. Birkhoff: *Dynamical Systems* (American Mathematical Society, Providence 1927)
32. D.O. Richstone, ApJ, **252**, 496 (1982)
33. L. Aguilar, P. Hut, J.P. Ostriker, ApJ, **335**, 720 (1988)

# Chaos and Chaotic Phase Mixing in Galaxy Evolution and Charged Particle Beams

Henry E. Kandrup

University of Florida, Gainesville, FL 32611, USA

**Abstract.** This paper discusses three new issues that necessarily arise in realistic attempts to apply nonlinear dynamics to galaxy evolution, namely: (i) the meaning of chaos in many-body systems, (ii) the time-dependence of the bulk potential, which can trigger intervals of *transient chaos*, and (iii) the self-consistent nature of any bulk chaos, which is generated by the bodies themselves, rather than imposed externally. Simulations and theory both suggest strongly that the physical processes associated with galactic evolution should also act in nonneutral plasmas and charged particle beams. This in turn suggests the possibility of testing this physics in real laboratory experiments, an undertaking currently underway.

## 1 Introduction

As recently as 1990, most galactic dynamicists ignored completely the possible role of chaos in galaxies. However, the past decade has seen a growing recognition that chaos can be important in determining the structure of real galaxies. Still, much recent work involving chaos in galactic astronomy has been simplistic in that it has involved naive applications of standard techniques from nonlinear dynamics developed to analyse two- and three-degree-of-freedom time-independent Hamiltonian systems. The object here is to discuss some of the additional complications which arise if nonlinear dynamics is to be applied to real galaxies, which are *many-body systems* comprised of a large number of interacting masses and characterised by a *self-consistently determined bulk potential* which, during their most interesting phases, can be *strongly time-dependent.*

## 2 Transient Chaos and Collisionless Relaxation

### 2.1 Transient Chaos Induced by Parametric Resonance

It is well known to nonlinear dynamicists that the introduction of an oscillatory time-dependence into even an otherwise integrable potential can trigger an interval of *transient chaos*, during which many orbits exhibit an exponentially sensitive dependence on initial conditions. Physically, this transient chaos arises from a resonance overlap between the frequencies $\sim \Omega$ of the unperturbed orbits and the frequency or frequencies $\sim \omega$ of the time-dependent perturbation.

In the past, the possible effects of such chaos have been considered for both nonneutral plasmas [1] and charged particle beams [2]. More recently, such transient chaos has also begun to be considered in the context of galactic astronomy [3]. That work has shown that, for large fractional amplitudes, $> 0.1$ or

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 154–168, 2003.

so, this resonance can be very broad, triggering significant amounts of chaos for $10^{-1} \leq \omega/\Omega \leq 10$; and that the existence of the phenomenon is robust, comparatively insensitive to details. It will, for example, persist if one allows for damped oscillations and/or modest drifts in frequency (generated, *e.g.,* by making $\omega$ a random variable sampling an Ornstein-Uhlenbeck colored noise process).

The breadth of the resonance and the insensitivity to details suggest that transient chaos could well prove common, if not generic, in violent relaxation [4], the collective process whereby a (nearly) collisionless galaxy or galactic halo evolves towards an equilibrium or near-equilibrium state. Violent relaxation typically involves damped oscillations triggered, *e.g.,* by interactions with another galaxy or, in the early Universe, by the cosmological details preceding galaxy formation. However, when considering collective effects there is only one natural time scale, the dynamical time $t_D \sim 1/\sqrt{G\rho}$, with $\rho$ a typical mass density, which determines both the characteristic orbital time scale and (at least initially) the oscillation time scale. The exact numerical values of these time scales will involve numerical coefficients which will in general be unequal and vary as a function of location within the galaxy. If, however, one need only demand that the oscillation and orbital time scales agree to within an order of magnitude, it would seem likely that this resonance could trigger transient chaos through large parts of the galaxy. In real galaxies the oscillations will presumably damp and the frequencies drift as the density changes and, presumably, power cascades from longer to shorter scales. To the extent, however, that the details are unimportant such variations should not obviate the basic effect.

### 2.2 Chaotic Phase Mixing and Collisionless Relaxation

But why might such transient chaos prove important in galactic evolution? Detailed numerical simulations indicate that violent relaxation can be a very rapid and efficient process, but simple models involving regular orbits, such as Lynden-Bell's [4] balls rolling in a pig-trough, do not approach an equilibrium nearly fast enough [5]. The important point, however, is that allowing for the effects of chaos can in principle dramatically accelerate both the speed and efficacy of violent relaxation. An initially localised ensemble of regular orbits evolved into the future in a time-independent potential will begin by diverging as a power law in time and, only after a very long period, slowly evolve towards a time-averaged equilibrium, *i.e.,* a uniform population of the *KAM* tori to which it is restricted. By contrast, a corresponding ensemble of chaotic orbits will begin by diverging exponentially at a rate that is comparable to a typical value of the largest finite time Lyapunov exponent for the orbits in the ensemble; and then converge exponentially towards an equilibrium or near-equilibrium state at a somewhat smaller, but still comparable, rate. The exponential character of this *chaotic phase mixing* means that the time scale associated with this process is typically far shorter than the time scale associated with *regular phase mixing* [6–8].

It is evident that chaotic phase mixing in a time-independent Hamiltonian system can trigger a very rapid approach towards an equilibrium, but this does not necessarily 'explain' violent relaxation. If, *e.g.,* most of the orbits in the system are regular, it would seem unlikely that chaotic phase mixing could be sufficiently ubiquitous to explain an approach towards a (near-)equilibrium for the galaxy as a whole. Indeed, one would expect that, for a galaxy that is in or near equilibrium the relative measure of chaotic orbits should be comparatively small: If the galaxy exhibits nontrivial structures like a bar or a cusp, the types of structures which one has come to associate with chaos, one would also expect large numbers of regular orbits must be present to serve as a 'skeleton' to support that structure [9]. Moreover, it is evident that, although chaotic mixing in a time-independent potential can be very efficient in mixing orbits on a constant energy surface, the energy of each particle remains conserved, so that there can be no mixing in energies. The extent to which chaotic phase mixing in a time-dependent potential will trigger an efficient shuffling of energies is not completely clear.

The important point, then, is that chaotic phase mixing associated with transient chaos in a time-dependent potential is likely to explain these remaining lacunae. At least for large amplitude perturbations, (say) 10% or more, this parametric resonance can trigger a huge increase in the relative abundance of chaotic orbits so that, for pulsation frequencies near the middle of the resonance, virtually all the orbits exhibit substantial exponential sensitivity. Moreover, given that this chaos involves a resonant coupling, it tends typically to cause a substantial shuffling of energies: those frequencies which are most apt to trigger lots of chaos are also apt to induce the largest shuffling of energies.

Still it should be noted that one *can* get a 'near-complete' shuffling of orbits on different constant energy surfaces even if the orbital energies are not especially well shuffled. This, however, is not necessarily a problem. Simulations of systems exhibiting efficient collective relaxation do not necessarily involve masses which completely 'forget' their initial conditions. Rather, comparatively efficient and complete violent relaxation is completely consistent with an evolution in which masses 'remember' (at least partially) their initial binding energies, *i.e.,* in which masses that start with comparatively large (small) binding energies end up with comparatively large (small) binding energies [10].

That it may be possible to achieve efficient chaotic phase mixing in an oscillating galactic potential while still relaxing towards a nearly integrable state within $10t_D$ or so is illustrated in Fig. 1. This Figure was generated from orbits evolved in a time-dependent potential of the form

$$V(x,y,z,t) = -\frac{m(t)}{(1+x^2+y^2+z^2)^{1/2}}, \qquad m(t) = 1+\delta m \frac{\sin \omega t}{(t_0+t)^p}, \qquad (1)$$

with $\delta m = 0.5$, $t_0 = 100$ and $p = 2$, which represents a galaxy damping towards an integrable Plummer sphere. The four curves in the top panel exhibit the $x$-component of the phase space *emittance*, $\epsilon_i = (\langle r_i^2\rangle\langle v_i^2\rangle - \langle r_i v_i\rangle^2)^{1/2}$ ($i = x, y, z$), all computed for the same localised ensemble of initial conditions, but allowing for four different frequencies $\omega$. The curves exhibit considerable structure but, at least for early times, the overall evolution is exponential. The bottom panels exhibit the $x$ and $y$ coordinates at five different times for the ensemble represented

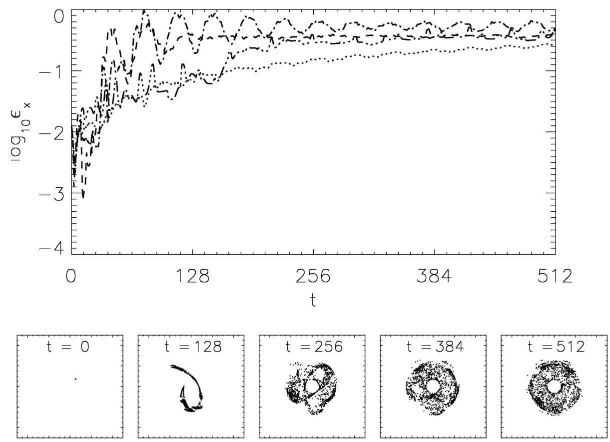



**Fig. 1.** (top) The emittance $\epsilon_x$ for an ensemble of initial conditions evolved in an integrable Plummer potential subjected to damped oscillations with four different frequencies: $\omega = 0.035$ (triple-dot-dashed), 1.40 (dashed), 3.50 (dot-dashed), and 7.00 (dotted). (bottom) $x$-$y$ scatter plots corresponding to the uppermost curve with $\omega = 3.50$.

by the uppermost of the four curves. Here $t_D \sim 20$, so that $t = 256$ corresponds to roughly $12t_D$.

Intuitively, one might expect that strong oscillations, which trigger the largest finite time Lyapunov exponents and the largest number of chaotic orbits, would yield the fastest chaotic phase mixing and, hence, the most rapid and most complete violent relaxation. A time-dependence with a weaker oscillatory component, *e.g.,* a time-dependence corresponding to a near-homologous collapse, might instead be expected to yield less chaos and, hence, less efficient and less complete violent relaxation. There is, therefore, an important need to determine the extent to which, in real simulations of violent relaxation, many/most of the orbits (or phase elements) are strongly chaotic, and the degree to which the rate and completeness of the observed violent relaxation correlate with the size of the largest finite time Lyapunov exponents and/or the relative measure of chaotic orbits. Investigations of these issues are currently underway.

## 3 The Role of Discreteness Effects

### 3.1 Microchaos and Macrochaos

The discussion in the preceding section, like most applications of nonlinear dynamics to galactic astronomy, neglects completely discreteness effects associated with the 'true' many-body potential, assuming that masses in a galaxy can be approximated as evolving in a smooth, albeit time-dependent, three-dimensional potential and that 'chaos' has its usual meaning. That this is justified is not completely obvious. The gravitational $N$-body problem for a large number of bodies of comparable mass is strongly chaotic in the sense that individual orbits have large positive Lyapunov exponent $\chi_N$ *even when there is absolutely no chaos in the continuum limit!* If, *e.g.,* a smooth density distribution corresponding to an integrable potential is sampled to generate an $N$-body density distribution, one finds that orbits evolved in this $N$-body distribution will be strongly chaotic, even for very large $N$, despite the fact that characteristics in the smooth potential generated from the same initial condition are completely integrable. Moreover, there is no sense in which the exponential sensitivity decreases with increasing $N$: if anything $\chi_N$ is an increasing function of $N$ [11]. In this sense, *larger N implies more chaos, not less!*

This situation has led some astrophysicists to question, either implicitly or explicitly, the reliability of the entire smooth potential approximation. Thus, *e.g.,* it has been suggested [12] that "the approximation of a smooth potential is useful for studying orbits, but not for studying their divergence." This is of course a problematic statement in that the distinction between exponential and power law divergence, emblematic of the differences between regular and chaotic behaviour, lies at the heart of applications of nonlinear dynamics to galactic dynamics. If the Lyapunov exponents associated with the bulk potential have nothing to do with the $N$-body problem, one must perforce reject completely all conventional applications of nonlinear dynamics to galactic astronomy.

The crucial point, then, is that there *does* appear to be a well-defined continuum limit, even at the level of individual orbits [13–15]. Suppose that a smooth density distribution, corresponding either to an integrable potential or to a potential admitting large measures of regular orbits, is sampled to generate a fixed, *i.e.* frozen in time, $N$-body density distribution, and that the trajectories of test particles evolved in this frozen distribution are compared with smooth potential characteristics with the same initial conditions. In this case, there is a precise sense in which, as $N$ increases, the frozen-$N$ trajectories converge towards the smooth potential characteristics. Both visually and in terms of the *complexity* [16] of their Fourier spectra, the frozen-$N$ trajectories come to more closely resemble the smooth potential characteristics; and, *viewed mesoscopically,* the frozen-$N$ and smooth potential orbits remain closer in phase space for progressively longer times. In particular, a frozen-$N$ orbit corresponding to an integrable characteristic will have a large Lyapunov exponent $\chi_N$ even if, visually, it is essentially indistinguishable from the regular characteristic!

But how can this be? The key recognition here is that two 'types' of chaos can be present in the $N$-body problem, characterised by *two different sets of Lyapunov exponents associated with physics on different scales.* Close encounters between particles trigger *microchaos*, a generic feature of the $N$-body problem, which leads to large positive Lyapunov exponents $\chi_N$. If, however, the bulk smooth potential is chaotic, one will also observe *macrochaos*, which is again characterised by positive, albeit typically much smaller, Lyapunov exponents $\chi_S$. Suppose, for example, that one compares the evolution of two nearby chaotic initial conditions in a single frozen-$N$ background or the same chaotic initial condition evolved in two different frozen-$N$ realisations of the same bulk density. In this case, one typically observes a three-stage evolution, namely: (1) a rapid exponential divergence at a rate $\chi_N$ set by the true Lyapunov exponents associated with the $N$-body problem, which persists until the separation becomes large compared with a typical interparticle spacing; followed by (2) a slower exponential divergence at a rate comparable to the (typically much smaller) smooth potential Lyapunov exponent $\chi_S$, which persists until the separation becomes macroscopic; followed by (3) a power law divergence on a time scale $\propto (\ln N)t_D$. For regular initial conditions, the second stage is absent and the time scale for the third stage scales instead as $N^{1/2}t_D$.

Microchaos becomes stronger as $N$ increases in the sense that the value of $\chi_N$ increases with increasing $N$ [17]. Despite this, however, it becomes progressively less important macroscopically in that the *range* of the chaos, *i.e.,* the scale on which the microchaos-driven exponential divergence of nearby orbits terminates, decreases with increasing $N$. In the limit $N \to \infty$ microchaos will become completely irrelevant but, for finite $N$, it does have an effect, at least on sufficiently short scales; and it is possible from an $N$-body simulation to extract estimates of both $\chi_N$ and the typically much smaller $\chi_S$ [15].

### 3.2 Modeling Discreteness Effects as Friction and Noise

It has been long recognised that, for sufficiently small $N$ and/or over sufficiently long times, discreteness effects will not be completely negligible. Systems like galaxies are 'nearly collisionless' in the sense that the stars interact primarily via collective macroscopic forces associated with the bulk density distribution; but, at least in principle, if one waits long enough discreteness effects should have an appreciable effect.

Astronomers are accustomed to modeling discreteness effects in the context of a Fokker-Planck description analogous to that formulated originally in the context of plasma physics [18]. However, it is not completely clear to what extent this is really justified. The conventional Fokker-Planck description was formulated originally to extract statistical properties of orbit ensembles and distribution functions over long time scales, assuming implicitly that the bulk potential is regular. To what extent, then, can Langevin realisations of a Fokker-Planck equation yield reliable information about individual orbits over comparatively short time scales, particularly if the orbits are chaotic?

Analyses of flows in frozen-$N$ systems indicate [14] that, at the level of both orbit ensembles and individual orbits, discreteness effects can in fact be modeled *extremely* well by Gaussian white noise in the context of a Fokker-Planck description, allowing for a dimensionless diffusion constant $D \propto 1/N$, consistent with the predicted scaling $D \propto \ln \Lambda/N$, with $\Lambda$ the so-called Coulomb logarithm [18]. For localised ensembles of initial conditions corresponding to both regular and chaotic orbits, phase mixing in frozen-$N$ systems and phase mixing in smooth potentials perturbed by Gaussian white noise yield virtually identical behaviour, both in terms of the evolution of various phase space moments such as the emittance and the rate at which individual orbits in the ensemble exhibit nontrivial 'transitions', *e.g.,* passing through some *entropy barrier* from one phase space region to another. And similarly, a comparison of frozen-$N$ orbits and noisy smooth potential orbits with the same initial condition reveals that their Fourier spectra typically exhibit comparable complexities. Gaussian white noise is even successful in mimicking some of the effects of microchaos. If, *e.g.,* one tracks the divergence of two noisy orbits with the same chaotic initial condition evolved in a smooth potential, one observes the same three-stage evolution as for a pair of frozen-$N$ orbits evolved in two different frozen-$N$ potentials.

An example of this agreement is illustrated in Fig. 2, which exhibits data generated by averaging over 100 pairs of orbits evolved in frozen-$N$ density distributions which correspond in the continuum limit to a triaxial homogeneous ellipsoid with axis ratios 1.95 : 1.50 : 1.05, perturbed by a spherically symmetric central mass spikes ('black hole'). The top two solid curves represent (from top to bottom) results for $N = 10^{4.5}$ and $N = 10^{5.5}$. The four dotted curves represent analogous results derived for pairs of noisy orbits evolved from the same initial conditions in the smooth potential with (from top to bottom) diffusion constant $D = 10^{-4}$, $10^{-5}$, $10^{-6}$, and $10^{-7}$. The near-coincidence of the top two solid and dotted curves indicates that discreteness effects for $N = 10^{p+1/2}$ are well-mimicked by Gaussian white noise with $D = 10^{-p}$.

Such striking agreement suggests strongly that investigations of how orbits in smooth potentials are impacted by the introduction of friction and noise can provide important insights into the role of graininess in real galaxies. It is customary to assert that, in a system as large as a galaxy, discreteness effects reflecting close encounters between stars are unimportant because the relaxation

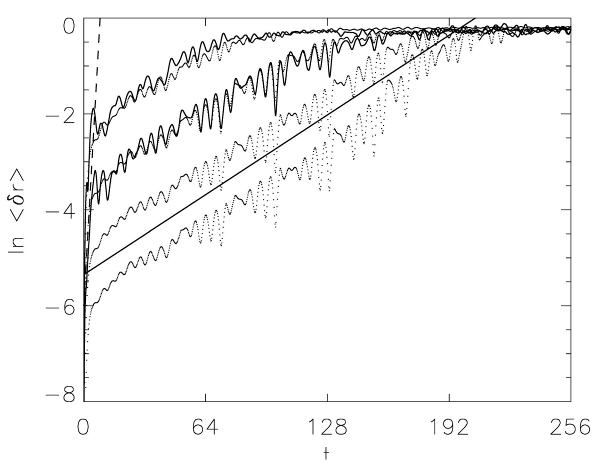



**Fig. 2.** The mean spatial separation between the same initial conditions evolved in two different frozen-$N$ backgrounds (solid curves) and different noisy orbits evolved in the smooth potential from the same initial condition (dots). The solid line has a slope 0.022, equal to the mean value of the smooth potential Lyapunov exponent $\chi_S$. The dashed curve has a slope 0.75, equal to the mean value of the $N$-body Lyapunov exponent $\chi_N$.

time $t_R$ on which they can induce appreciable changes in quantities like the energy is orders of magnitude longer than the age of the Universe [19]. This is likely to be true if the galaxy is an exact equilibrium, especially an equilibrium characterised by an integrable potential. However, the assertion is suspect if (as must usually be the case) the system is only 'close to' an equilibrium or near-equilibrium, especially if the bulk potential is characterised by a phase space admitting a complex coexistence of regular and chaotic orbits.

Over the past decade, analyses of flows in time-independent Hamiltonian systems have revealed that even very weak perturbations, idealised as friction and white noise corresponding to $t_R \sim 10^6 - 10^9 t_D$ and, hence, $D \sim 10^{-6} - 10^{-9}$, can have significant effects within a time as short as $100t_D$ or less by facilitating phase space diffusion through cantori or along the Arnold web [20–22]. The basic point is that the motions of chaotic orbits in a complex potential can be constrained significantly by topological obstructions like cantori or the Arnold web which, albeit not completely preventing motions from one phase space region to another, serve as an *entropy barrier* to impede such motions. In many respects, the physical picture is similar to the elementary problem of effusion of gas through a tiny hole in a wall. There is nothing in principle to prevent a gas molecule from passing through the hole and, hence, escaping from the region to which it is originally confined; but, if the hole is very small, the time scale associated with this effusion can be extremely long.

In the same sense, and for much the same reason, chaotic orbits trapped in one phase space region may, in the absence of perturbations, remain stuck in that region for a very long time. However, subjecting the orbits to noise will 'wiggle' them in such a fashion as to increase the rate at which they pass through the entropy barrier, thus accelerating phase space transport. Numerical simulations indicate that, in at least some cases, this escape process can be well approximated by a Poisson process, with the number of nonescapers decreasing exponentially at a rate $\Lambda$ that is determined by the perturbation [23, 24]. This effect appears to result from a resonant coupling between the orbits and the noise. White noise is characterised by a flat power spectrum and, as such, will couple to more or less anything. If, however, the noise is made coloured, *i.e.,* if instantaneous kicks are replaced by impulses of finite duration, the high frequency power is reduced; and, if the autocorrelation time becomes sufficiently long that there is little power at frequencies comparable to the orbital frequencies, the effect of the noise decreases significantly. Significantly, it appears that, overall, the details of the perturbation may be largely irrelevant: additive and multiplicative Gaussian noises tend to have comparable effects and the presence or absence of friction does not seem to matter. All that appears to matter are the amplitude and the autocorrelation time upon which there is a relatively weak, roughly logarithmic, dependence.

But what does all this imply for a real galaxy? Given that collisionless near-equilibria must be more common than true equilibria, it would seem quite possible that, during the early stages of evolution, a galaxy might settle down towards a near-equilibrium, rather than a true equilibrium, *e.g.,* involving what have been termed [25] ‘partially mixed’ building blocks. If discreteness effects and all other perturbative effects could be ignored, such a quasi-equilibrium might persist without exhibiting significant changes over the age of the Universe. If, however, one allows for discreteness effects or, alternatively, other perturbations reflecting, *e.g.,* a high density cluster environment, the orbits could become shuffled in such a fashion as to trigger significant changes in the phase space density and, consequently, a systematic secular evolution [26].

Such a scenario could, for example, result in the destabilisation of a bar. Many models of bars (*e.g.* [27]) incorporate ‘sticky’ [28] chaotic orbits as part of the skeleton of structure, replacing crucial regular orbits which can be absent near corotation and other resonances. Making these ‘sticky’ orbits become unstuck could cause the bar to dissipate. Similar effects could also cause an originally nonaxisymmetric cusp to evolve towards a more nearly axisymmetric state. To the extent that the triaxial Dehnen potentials are representative, one can argue that chaotic orbits may be extremely common near the centers of early-type galaxies, but that many of these chaotic orbits are extremely sticky [29] and, as such, could help support the nonaxisymmetric structure. Perturbations that make these sticky orbits wildly chaotic could *de facto* break the bones of the skeleton supporting the structure and trigger an evolution towards axisymmetry.

# 4 Experimental Tests of Galactic Dynamics

## 4.1 Similarities Between Galaxies and Nonneutral Plasmas

Even though electrostatics and Newtonian gravity both involve $1/r^2$ forces, electric neutrality implies that the physics of neutral plasmas is very different from the physics of self-gravitating systems. Viewed over time scales $> t_R$, nonneutral plasmas and charged-particle beams are also very different from self-gravitating systems: the attractive character of gravity leads to phenomena like evaporation and core-collapse which cannot arise in a beam or a plasma. If, however, one restricts attention to comparatively short times $\ll t_R$, much of the physics should be the same. Theoretical expectations, supported by numerical simulations, suggest that it is the existence of long range order, not the sign of the interaction, which is really important; but, to the extent that this be true, collisionless nonneutral plasmas and collisionless self-gravitating systems should be quite similar.

Typical sources of charged-particle beams configure the beams in trains of ‘packets’ or ‘bunches’, as they are termed by accelerator dynamicists. The objective of a good high-intensity accelerator is to generate bunches comprised of a large total number of charges confined to a small phase space volume and then accelerate those bunches to very high energies while minimising any growth in emittance. As one example, modern photocathode-based sources of electron beams routinely generate bunches comprised of some $10^{10} - 10^{11}$ electrons with transverse ‘emittance’ $\tilde{\epsilon}$ of a few microns. (Here $\tilde{\epsilon} = \epsilon/v_0$, where $v_0$ is the mean axial velocity of the particle distribution.) The energy relaxation time $t_R$ associated with such bunches typically corresponds to the time required for a bunch to travel a distance $\sim 1$ km or so which, in many cases, is much longer than any distance of experimental interest, so that the bunches are 'nearly collisionless.'

Models of equilibrium configurations of nonneutral plasmas and charged-particle beams confined by electromagnetic fields can be characterised by a complex phase space quite similar to that associated with models of elliptical galaxies and, as such, have orbits with very similar properties. For example [9], the so-called 'thermal equilibrium model' [32] of beam dynamics, which involves a self-interacting nonneutral plasma in thermal equilibrium confined by an anisotropic harmonic oscillator potential, is strikingly similar [29] to the non-spherical generalisations of the Dehnen potential of galactic astronomy in terms of such properties as the degree of 'stickiness' manifested by chaotic orbits or how the relative measure of chaotic orbits and the size of the largest Lyapunov exponent vary with shape.

As in galactic dynamics, questions have been raised regarding the validity of the continuum approximation for nearly collisionless charged particle beams [31]. However, comparatively short time integrations ($t \ll t_R$) involving discreteness effects and the nature of the continuum limit in nonneutral plasmas [33] yield results essentially identical to what is observed for gravity – although the behaviour associated with neutral plasmas is *very* different. In particular, the macroscopic manifestations of phase mixing, for both regular and chaotic orbits, are indistinguishable, and the coexistence of microchaos and macrochaos persists unabated.

Nontrivial effects associated with a time-dependent potential have also been predicted for both nonneutral plasmas [1] and charged particle beams [2]. Although the time-dependence that is envisioned in a beam is typically less violent than that anticipated in violent relaxation within a galaxy, such is not always the case. Indeed, there is compelling experimental evidence that, in a beam, such a time-dependence can have the undesireable effect of ejecting particles from the core into an outerlying halo [34].

Perhaps most interesting, however, is the fact that numerical simulations that reproduce successfully 'anomalous relaxation' observed in real laboratory experiments involving accelerator beams have shown compelling evidence of chaotic phase mixing. One classic example involves the propagation of five nonrelativistic high-intensity beamlets in a periodic solenoidal transport channel, where self-consistent space-charge forces are extremely important [35]. Ideally, these beamlets should exhibit coherent periodic oscillations (quite literally disappearing and reappearing) which might be expected to decay only on a relaxation time scale $t_R$ that corresponds to a propagation distance $\sim 1$ km. However, regardless of how well the beam was matched to the transport channel, the beamlets were seen to reappear only once, at a point $\sim 1$ m from the source, disappearing completely within 2 m or so (*cf.* Fig. 6.10 in [35]). Their failure to reappear again would seem to reflect some collisionless process that, in effect, causes the particles to 'forget' their initial conditions.

Detailed simulations using the particle-in-cell code *WARP* [36], which do an extremely good job of reproducing what is actually seen, demonstrate seemingly unambiguously that, because of the time-dependent space-charge potential, a

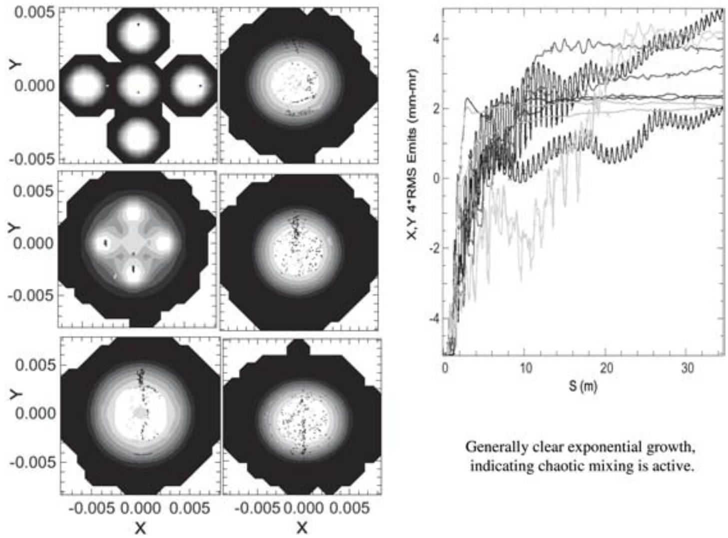



**Fig. 3.** Evolution of five representative ensembles of test particles in the five-beamlet simulation. The left hand panel shows snapshots of the ensembles at (top-to-bottom left column) 0 m, 0.98 m, 2.88 m, and (top-to-bottom right column) 5.24 m, 11.52 m, and 31.68 m, with the $x$- and $y$-axes labeled in meters. The right panel shows the evolution of the logarithm of the emittances $\epsilon_x$ and $\epsilon_y$ as a function of distance $S(z)$ along the accelerator.

large fraction of the particles in the beam experience the effects of strong, possible transient, macrochaos [37, 38]. This is, *e.g.,* evident from Fig. 3, which illustrates the evolution of representative test particles which interact with the bulk potential but not with each other. Here the left hand panel shows snapshots of the beam after it has travelled distances 0 m, 0.98 m, 2.88 m, 5.24 m, 11.52 m, and 31.68 m, with the representative ensembles superimposed. The right hand panel exhibits the evolution of the emittances $\epsilon_x$ and $\epsilon_y$ for these ensembles. It is evident that the initially localised ensembles are diverging exponentially so as to fill much of the accessible phase space, and that this exponential divergence coincides with the beamlets losing their individual identities. Also evident is the fact that the behaviour observed here is very similar to that exhibited in Fig. 1 which, recall, was generated for orbits in a perturbed Plummer potential exhibiting damped oscillations.

### 4.2 Testing Galaxy Evolution with Charged Particle Beams

The aforementioned similarities between galactic astronomy and charged particle beams suggest the possibility of using accelerators as a laboratory for astrophysics in which one can perform experimental tests of galactic dynamics, a possibility currently being developed by a University of Florida – Fermilab/Northern Illinois University – University of Maryland collaboration. This collaboration, which has the dual aims of (1) obtaining an improved understanding of the applicability of nonlinear dynamics to nearly collisionless systems interacting via long range forces and (2) using that understanding to generate more sharply focused bunches by minimising undesirable increases in emittance, is currently planning concrete experiments which can, and presumably will, be performed on the University of Maryland Electron Ring (*UMER*) currently under construction. Here a number of obvious issues, all experimentally testable, come to mind:

How ubiquitous is chaotic phase mixing as a source of anomalous relaxation? Older experiments with lower-intensity beams, where the space-charge forces were comparatively unimportant, tended not to manifest extreme examples of anomalous relaxation. Anomalous relaxation appears more common in high intensity beams, especially in settings where a time-dependent density distribution generates a strongly time-dependent potential; and it is obvious to ask whether chaos is the principal culprit. The idea here is to identify the types of scenaria that tend generically to yield anomalous relaxation and to determine, *e.g.,* whether such scenaria tend typically to be associated with a bulk potential that incorporates a strong, roughly oscillatory component. Do numerical simulations of orbits ensembles evolved in such systems exhibit evidence of chaotic phase mixing? And do individual orbits in those ensembles exhibit strong exponential sensitivity, associated, *e.g.,* with transient chaos?

Do instabilities tend to trigger transient chaos? Instabilities in collisionless systems can exhibit behaviour qualitatively similar to that associated with turbulence in collision-dominated systems, but it is well known that turbulence is a strongly chaotic phenomenon. This possibility is especially interesting in that turbulence is another setting where different 'types' of chaos, characterised by wildly different time scales, can act on different length scales.

What types of geometries, both strongly time-dependent and nearly time-independent, tend to yield the most efficient chaotic phase mixing and the largest measures of chaotic orbits? Do time-dependent evolutions involving strongly convulsive oscillations tend generically to exhibit especially fast relaxation? And do they tend to yield especially large amounts of chaos, as probed by the relative measure of chaotic orbits and/or the sizes of the largest (finite time) Lyapunov exponents? To the extent that bulk properties of such 'accelerator violent relaxation' correlate with the degree of chaos exhibited in the evolving beam, and that the degree of chaos correlates with the form of the macroscopic time dependence, one will have a physically well-motivated explanation of which sorts of scenaria would be expected to exhibit complete and efficient violent relaxation and which would not!

Is it, *e.g.,* true that, for nearly axisymmetric configurations, prolate (or oblate) bunches tend generically to exhibit especially large amounts of chaos? And do any such trends that are observed coincide with trends observed in models of galactic equilibria [29]? Even if a beam bunch remains nearly axisymmetric during its evolution, the acceleration mechanism can – and in general will – change its shape as it passes down an accelerator in a fashion that depends on the accelerator design. The obvious question, then, is whether the oblate or prolate phase tends generically to be especially chaotic.

Addressing this and related issues could provide important insights as to *why* galaxies have the detailed shapes that they do, a general question for which, at the present time, no compelling dynamical explanation exists. One knows, *e.g.*, that elliptical galaxies tend to have isophotes that are slightly boxy or disky, and that this boxiness or diskiness correlates with such properties as the rotation rate, the steepness of the central cusp, and the size of any deviations from axisymmetry [39]. Must all these effects be attributed to the detailed form of the formation scenario, or is there a clear dynamical explanation? Is it, *e.g.,* true that the observed deviations from perfect ellipsoidal symmetry conspire to reduce the relative number of chaotic orbits or to increase the numbers of certain regular orbit types required as a skeleton to support the observed structure?

One might also use accelerator experiments to probe the role of discrete substructures and the extent to which they can be modeled as friction and noise in the context of a Fokker-Planck description [40]. If the injection of a beam involves a large mismatch, a significant charge redistribution will occur, resulting in violent relaxation, 'turbulent' behaviour, and the formation of substructures ('lumps') on a variety of scales. To the extent that such a time-dependent evolution can be described in a continuum approximation, one might then expect that the bulk potential will correspond to a highly complex time-dependent phase space and that the substructures could act as a 'noisy' source of extrinsic diffusion, facilitating both transitions between 'sticky' and 'wildly chaotic' behaviour and, in some cases, transitions between regularity and chaos. Given the evidence (*cf.* [33]) that, at least over short times, discreteness effects act similarly for attractive and repulsive $1/r^2$ forces, such insights could be directly related to such issues as the destabilisation of bars in spirals and/or the secular evolution of nonaxisymmetric ellipticals towards more nearly axisymmetric states.

Do systems tend to evolve in such a fashion as to minimise the amount of chaos? There is an intuitive expectation amongst many galactic astronomers (*cf.* [41]) that galaxies tend to evolve towards equilibria which incorporate few if any chaotic orbits, *i.e.,* that nature somehow favors nearly-regular equilibria. It would certainly appear true that a model must incorporate significant numbers of regular orbits to support interesting structures like bars and/or triaxiality, but this does not *a priori* preclude the possibility of chaotic orbits also being present. Generic time-independent three-degree-of-freedom potentials are neither completely regular nor completely chaotic, admitting instead a complex coexistence of regular and chaotic phase space regions. The obvious question then is: are galactic equilibria or near-equilibria typically well-represented by potentials which are generic in this sense; or are they, for reasons unknown, special in that they tend to be rather nearly regular?

## 5 Conclusions

This paper has focused on several fundamental issues that arise in attempts to apply nonlinear dynamics to real galaxies, many-body systems characterised by a self-consistently determined bulk potential which, during their most interesting phases, can be strongly time-dependent. As recently as a decade ago these issues would have been considered of largely academic, rather than practical, interest. However, recent observational advances – which facilitate improved high resolution photometry of individual objects as well as statistical analyses of large samples with varying redshift – and improved computational resources – which allow unparalleled explorations of multi-scale structure –, together with the recognition that the basic physics can be also probed in the context of charged particle beams, imply that theoretical predictions regarding these 'academic' issues can in fact be tested observationally, numerically, and experimentally.

Thanks to Ioannis Sideris for providing Fig. 1 and to Rami Kishek for providing Fig. 3. Thanks also to Court Bohn for his comments on a preliminary draft of the manuscript. Partial financial support was provided by NSF AST-0070809.

## References

1. S. Strassburg and R. C. Davidson: Phys. Rev. **E61**, 5753 (2000)
2. R. Gluckstern: Phys. Rev. Lett **73**, 1247 (1994)
3. H. E. Kandrup, I. M. Vass, and I. V. Sideris: Mon. Not. R. astr. Soc.: submitted (2002) (astro-ph/0211056)
4. D. Lynden-Bell: Mon. Not. R. astr. Soc. **136**, 101 (1967)
5. H. E. Kandrup: 'Collisionless Relaxation of Stellar Systems'. In: *Galaxy Dynamics, A Rutgers Symposium*, ed. by D. Merritt, J. N. Sellwood, and M. Valluri (Sheridan, San Francisco 1999)
6. H. E. Kandrup and M. E. Mahon: Phys. Rev. **E49**, 3735 (1994)
7. M. E. Mahon, R. A. Abernathy, B. O. Bradley, and H. E. Kandrup: Mon. Not. R. astr. Soc. **275**, 443 (1995)
8. D. Merritt and M. Valluri, Astrophys. J. **471**, 82 (1996)
9. J. Binney: Comments Astrophys. **8**, 27 (1978)
10. P. J. Quinn and W. H. Zurek: Astrophys. J. **331**, 1 (1988)
11. M. Hemsendorff and D. Merritt: astro-ph/0205538 (2002)
12. D. Heggie: 'Chaos in the $N$-Body Problem of Stellar Dynamics'. In: *Predictability, Stability, and Chaos in N-Body Dynamical Systems*, ed. A. E. Roy (Plenum, New York 1991)
13. H. E. Kandrup and I. V. Sideris: Phys. Rev. **E64**, 056209-1 (2001)
14. I. V. Sideris and H. E. Kandrup: Phys. Rev. **E65**, 066203-1 (2002)
15. H. E. Kandrup and I. V. Sideris: Astrophys. J. in press (2003) (astro-ph/0207090)
16. H. E. Kandrup, B. L. Eckstein, and B. O. Bradley: Astron. Astrophys. **320**, 65 (1997)
17. I. V. Pogorelov: Phase Space Transport and the Continuum Limit in Nonlinear Hamiltonian Systems. Ph. D. Thesis, University of Florida, Gainesville (2000)
18. M. N. Rosenbluth, W. M. MacDonald, and D. L. Judd: Phys. Rev. **107**, 1 (1957)
19. J. Binney and S. Tremaine, *Galactic Dynamics* (Princeton University, Princeton, 1987)

20. H. E. Kandrup: Astrophys. J. **480**, 155 (1997)
21. C. Siopis and H. E. Kandrup: Mon. Not. R. astr. Soc. **391**, 43 (2000)
22. H. E. Kandrup and S. J. Novotny: Celestial Mechanics, submitted (2002) (astro-ph/0204019)
23. I. V. Pogorelov and H. E. Kandrup: Phys. Rev. **E60**, 1567 (1999).
24. H. E. Kandrup, I. V. Pogorelov, and I. V. Sideris: Mon. Not. R. astr. Soc. **311**, 719 (2000)
25. D. Merritt and T. Fridman: Astrophys. J. **460**, 136 (1996)
26. H. E. Kandrup: Space Science Reviews, in press (2003) (astro-ph/0011302)
27. P. A. Patsis, E. Athanassoula, and A. C. Quillen: Astrophys. J. **483**, 731 (1997)
28. G. Contopoulos: Astron. J. **76**, 147 (1971)
29. H. E. Kandrup and C. Siopis: Mon. Not. R. astr. Soc.: submitted (2002)
9. C. L. Bohn and I. V. Sideris: Phys. Rev. **ST AB**: submitted (2002)
31. J. Struckmeier, Phys. Rev. **E54**, 830 (1996)
32. M. Brown and M. Reiser: Phys. Plasmas **2**, 965 (1995)
33. H. E. Kandrup, I. V. Sideris, and C. L. Bohn: Phys. Fluids, submitted (2002)
34. R. W. Garnett *et al*: In *Linac 2002: Proceedings of the XXI International Linac Conference*, in press.
35. M. Reiser: *Theory and Design of Charged Particle Beams* (Wiley, New York, 1994)
36. D. P. Grote, A. Friedman, I. Haber, and S. Yu: Fusion Eng, Design **32-33**, 193 (1996)
37. R. A. Kishek, P. G. O'Shea, and M. Reiser: Phys. Rev. Lett. **85**, 4514 (2000)
38. R. A. Kishek, C. L. Bohn, P. G. O'Shea, M. Reiser, and H. E. Kandrup: In: *Proceedings of the IEEE 2001 Particle Accelerator Conference*, ed P. Lucas and S. Weber (IEEE, Chicago, 2001)
39. J. Kormendy and R. Bender: Astrophys. J. Lett. **464**, 119 (1996)
40. C. L. Bohn and J. R. Delayen: Phys. Rev. **E50**, 1516 (1994)
41. M. Schwarzschild: Astrophys. J. **232**, 236 (1979)

xx

# Motions of a Black Hole near the Center of a Galaxy

Richard H. Miller

University of Chicago, Astronomy Department,
5640 S. Ellis Ave., Chicago 60637-1433, U. S. A.

**Abstract.** Some years ago we published an account of experiments which indicated that the nucleus of a galaxy orbits around the mass centroid. This can be viewed as an orbiting density wave which grows near the center in a galaxy model that starts without such motions. While these experiments were run without a massive particle, we suggested that similar physical effects might cause a massive particle near the center to oscillate with larger amplitudes than indicated by simple Brownian motion arguments. Results from recent experiments will be reported to clarify some of the issues raised by a massive particle (a black hole) near the center.

Motions of a massive black hole near the center of a galaxy have excited considerable interest lately, since a typical galaxy may harbor a massive black hole near its center [1]. More recent work is described in a recent review by Ferrarese and Merritt [2].

Dynamical effects cause a cusp to build up around a black home. A few of the obvious questions follow:
• the amount of mass in the cusp,
• the density profile of the cusp,
• black hole motions within the cusp, and
• larger scale motions as the black hole–cusp combination move together.

Our methods are best suited to the larger scale motions of the combined black hole and cusp, and those aspects are the subject of this paper. The other questions require a different computational approach. Ferrarese and Merritt discuss some of these features.

Our approach is experimental. It is based on numerical experiments carried out in a computer, using $n-$body programs. Results to be presented in this note are best understood with some appreciation of the methods used to obtain them.

## 1 Experimental Details

The $n-$body approach produces an initial value problem, which involves one set of computer programs to establish the starting condition (the loader) and another set to advance the system in time (the integrator). Of course, the state of the system at any stage along the way can be regarded as a new initial condition.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 169–184, 2003.

## 1.1 Generalities

We refer to the black hole throughout this paper simply as a "massive particle," abbreviated "MP," to stress the fact that it is being treated as if it were a simple Newtonian particle. Unusual features of a black hole caused by general–relativistic effects are confined to spatial regions which are much too small to resolve in many $n-$body studies, including this one. Field particles feel the black hole only as if it were a massive particle acting through Newtonian gravitation. Forces due to the MP are softened to avoid excessive forces.

### Boundary Conditions

Boundary conditions are important aspects of any self–consistent self–gravitating problem like that considered here. Periodic boundary conditions provide a way to avoid having to follow the dynamics of an entire galaxy. The idea is to provide a suitable representation of the region surrounding the MP and its associated cusp that at once is computationally manageable and provides an appropriate physical surrounding. The immediate environs of the MP are also treated as being isotropic.

### Equipartition

The notion of "equipartition" is often used in this note. It comes from statistical mechanics, where it refers to equal mean energies per degree of freedom, and it is used in that sense. There is no reason to believe that equipartition should hold under the present circumstances, but the concept provides a convenient terminology to discuss the phenomenon that MP motions do not quite fit the expected pattern.

A workaround is to define an "effective" MP mass. Call the "actual" MP mass $M_{\rm act}$ and the "effective" mass $M_{\rm eff}$, where the mean MP energy would be the same as the mean field particle energy if the MP mass were $M_{\rm eff}$. The relation between $M_{\rm eff}$ and $M_{\rm act}$ is of interest.

But a warning is called for: constant reference to $M_{\rm eff}$ tends to suggest that there is some "mass" that so moves, while the physical reality is likely to be a more abstract concept like a "density wave" or a "mode." Certainly there is no identifiable "thing" associated with the excess of $M_{\rm eff}$ over $M_{\rm act}$.

## 1.2 Loader

An isothermal configuration is generated by the loader. Some design considerations follow.

### Isothermal Equation Around a Massive Particle

The standard recipe for generating an isothermal is to use an isothermal distribution function [19], $f = f(\mathcal{E}) = \mathcal{N} \exp(-\mathcal{E}/E_0)$, where $\mathcal{E}$ is the energy of a particle moving within the system, $E_0$ is the (one dimensional) velocity dispersion of field particles, and $\mathcal{N}$ is a normalization factor. Use energies per unit mass, so $\mathcal{E} = \frac{1}{2}v^2 + U_{\text{tot}}$, where $U_{\text{tot}}$ is the *total* potential within which the particle moves. For present purposes, $U_{\text{tot}}$ is made up of three parts, which happily add linearly. Those parts are (1) $U_{\text{sc}}$, the self–consistent potential generated by the complexion of field particles, (2) $U_\bullet$, the potential around the massive particle, and (3) $U_{\text{ext}}$ an external potential. The external potentials that will be considered are (isotropic) harmonic, and the potential due to the MP is Keplerian.

Densities are computed in the usual way by integrating over velocities at the selected point in configuration space.

$$\rho = \int f d^3v = \rho_0 \exp\left(-\frac{U_{\text{tot}}}{E_0}\right) = \rho_0 \exp\left(-\frac{U_{\text{sc}} + U_\bullet + U_{\text{ext}}}{E_0}\right), \tag{1}$$

where constants from the integration and the normalization factor have been absorbed into a single constant, $\rho_0$.

**Remark.** An analytic self–consistency problem can be formulated for spherical symmetry by incorporating this density into the Poisson equation to yield

$$\frac{1}{r^2}\frac{d}{dr}\left(r^2\frac{dU_{\text{sc}}}{dr}\right) = 4\pi G\rho_0 \exp\left(-\frac{U_\bullet + U_{\text{ext}}}{E_0}\right)\exp\left(-\frac{U_{\text{sc}}}{E_0}\right). \tag{2}$$

Save for the first exponential, this is the standard isothermal equation [4]. Polytropic systems follow the same line of argument, with $f_{\text{poly}} = \mathcal{N}_{\text{poly}}(-\mathcal{E}/E_0)^p$. Because of the extra terms resulting from $U_\bullet$, neither the polytropic nor the isothermal cases have the usual homology invariances [4].

The singularity in $U_\bullet$ as $r \to 0$ requires special treatment in the integration of (2). Integrating once, (2) becomes

$$r^2\frac{dU_{\text{sc}}}{dr} = 4\pi G\rho_0 \int_{r0}^{r} \exp\left(-\frac{U_\bullet + U_{\text{sc}}}{E_0}\right) r^2\, dr,$$

if $U_{\text{ext}}$ is set to zero. Huntley and Saslaw [5, 6] discuss treatment of the lower limit, $r0$, on the integral to work around the singularity in $U_\bullet$. The contribution of $U_\bullet$ is more complicated in stellar dynamical problems than they indicate.

## A Practical Loader

Fortunately, computational solutions to the Poisson equation usually work from the integral form and get around issues of singular potentials by softening the forces. In the absence of analytic solutions to the self–consistency problem in periodic boundary conditions, we use iterative solutions built with densities in the form provided by (1).

The iterative procedure runs as follows. Guess a density distribution, then find the Newtonian potential generated by that density distribution using the same Poisson solver as for the integrations. Add any external potential and the potential due to the massive particle, and then take the exponential of that summed potential at every tabulated point to generate a new density distribution. Taking that new density distribution as a new starting point, repeat the process until it converges in some sense. Happily, this process converges nicely.

Care is required on a few points: to obtain the desired number of particles, to determine the final value for the constant that replaces the physical gravitational constant, to obtain the desired ratio of maximum to minimum densities, to handle scaling and additive constants in the three kinds of potential consistently, and so on.

This builds a density from which each field particle can be assigned a position by a quasi–Monte Carlo process. Once given a location, each particle is assigned a velocity sampled from an isotropic Maxwellian distribution with $E_0$ set to ensure self–consistency. The MP was always loaded on the origin.

Loads so generated are usually very near self–consistency, as tested by integrating them forward in time and noting that most properties are constant to within a percent or so. Polytropic models differ only in detail, but none were used in the experiments discussed here.

### 1.3 Integrator

A time–centered leapfrog integrator is used. It is symplectic, and so guarantees a Liouville theorem, which is an essential ingredient of the physical properties of a stellar dynamical system. Particles can cross a “periodic boundary” freely.

Forces are derived from a potential, which is worked out anew at each integration step. The self–consistent part is computed a Poisson solver, and any external potential is then added. Forces due to the massive particle are handled in the integrator to avoid having to treat large forces in the Poisson solver. Tabulated densities serve as input to the Poisson solver, and the output is tabulated at the same set of points. Our set of tabulation points forms a cubic cartesian lattice with lattice spacing $L$. The periodicity length is $N$ tabulation points, so a the edge of a periodic cell is $NL$.

Field particles have mass, $m$, and $T$ is our integration time step. Physically consistent units are related by the dimensionless constant $W = GmT^2/L^3$, which replaces the physical gravitational constant in the calculations. Numerical values quoted later in this paper are given in dimensionless form, so distances are in units of $L$, masses in units of $m$, and velocities in units of $L/T$, and so on. MP masses are quoted in units of $m$, so an MP mass of 100K is $100\,000m$. There is no simple way to relate these units to observations.

## 2 Results from Experiments

Experiments reported in this paper came from 3 sets: (1) Periodicity length $N = 128$, (2) Periodicity length $N = 64$, and (3) the entire configuration embedded in a strong isotropic harmonic potential, again with periodicity length $N = 128$. All experiments ran with $P = 800\,768$ particles. The MP was initially at the center of a periodic cell, so the even values of $N$ used here place the MP at equal distances from eight neighboring tabulation points, at the common corner of the small cubic cells centered on those 8 tabulation points. Even very small motions cause the MP to move through various ones of those small cells.

Experiments with Periodicity Length $N = 128$ are described first, followed by the remaining sets, which begin with Sect. 2.4.

## 2.1 Nature of MP Motions

MP motions look like a growing oscillation which levels off at later times. The $Y-$component of the MP position is shown as a function of time for six runs in Fig. 1, which are stacked for clarity. The numbers at the end of each track indicate the mass of the MP. Amplitudes for the 100K track are magnified 5 times and they are doubled for the 10K track.

While the tracks at lower mass ($< 10$K) look like a growing oscillation that levels off at some amplitude, all with nearly the same period, tracks for the more massive 10K and 100K experiments look different. Both the amplitudes and periods are irregular. It is tempting to guess that they may become chaotic, but we cannot make any definitive claims in that regard.

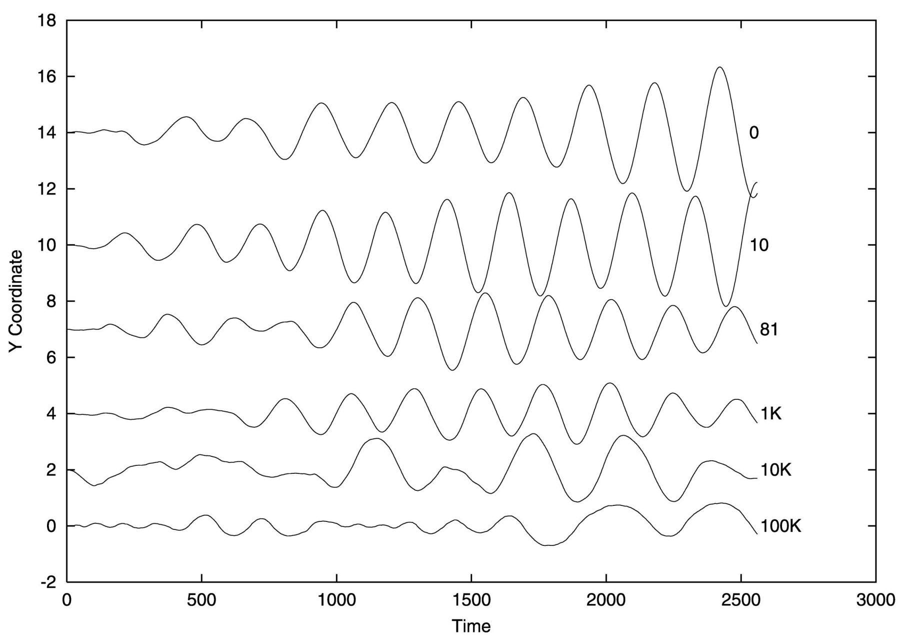



**Fig. 1.** Y-Component of MP motions from Experiments with Different MP Masses. Tracks are spaced vertically for clarity. Labels at the end of each track indicate the mass of the MP

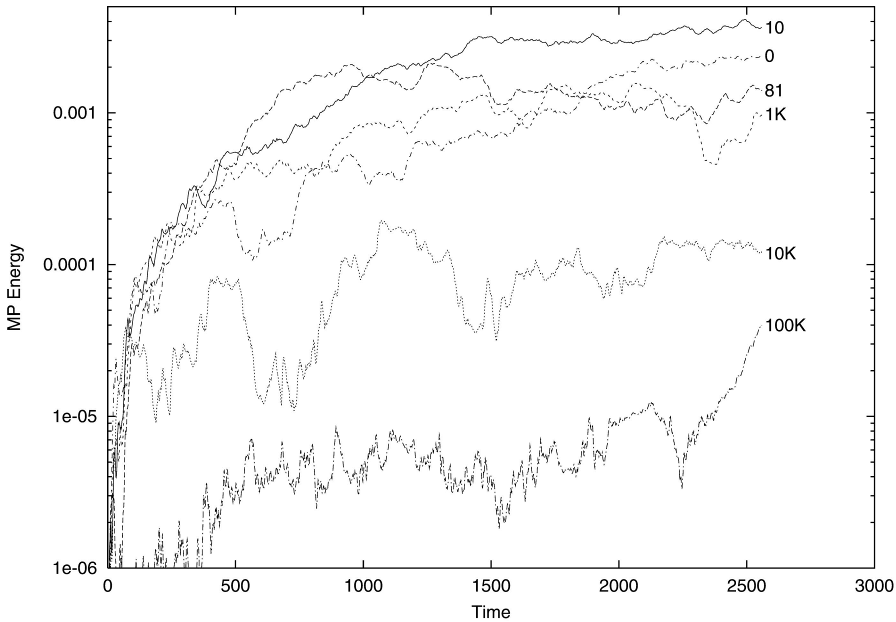



**Fig. 2.** Energies per Unit Mass from Experiments with Different MP Masses

## 2.2 MP Energies and Effective Masses

Orbital periods are sufficiently constant throughout these runs that a total MP energy per unit mass can be estimated from $\frac{1}{2}((\omega x)^2 + v^2)$. A convenient way to estimate $1/\omega$ is to scale the displacements so they match the velocities throughout the runs, which they do pretty well. When energies so determined are plotted as functions of time, all on the same page, the Fig. 2 is generated.

Log (energy per unit mass) is plotted in Fig. 2. At smaller MP masses the curves nearly overlap, but for the greatest MP masses, they "equilibrate" at lower values. The near overlap is of interest here.

The track for MP mass of 100K climbs rapidly at the end. The MP had wandered off to a large distance on a nice smooth orbit.

Tracks in Fig. 2 that are associated with runs with smaller MP masses are reasonably smooth, while those with the largest MP masses become increasingly irregular, even apart from the high–frequency "grass" along the track for 100K. Part of that "grass" arises from roundoff.

### Weight Energies by "Effective" MP Mass

Tracks can each be scaled so that they all settle down at about the 0.75 level, the mean energy of a field particle, giving Fig. 3. In the picture provided by equipartition, the scale factor represents an "effective" mass.

The overlap is by no means exact. The irregularity in a given track may be regarded as typical of the scatter of MP energy values due to the statistical

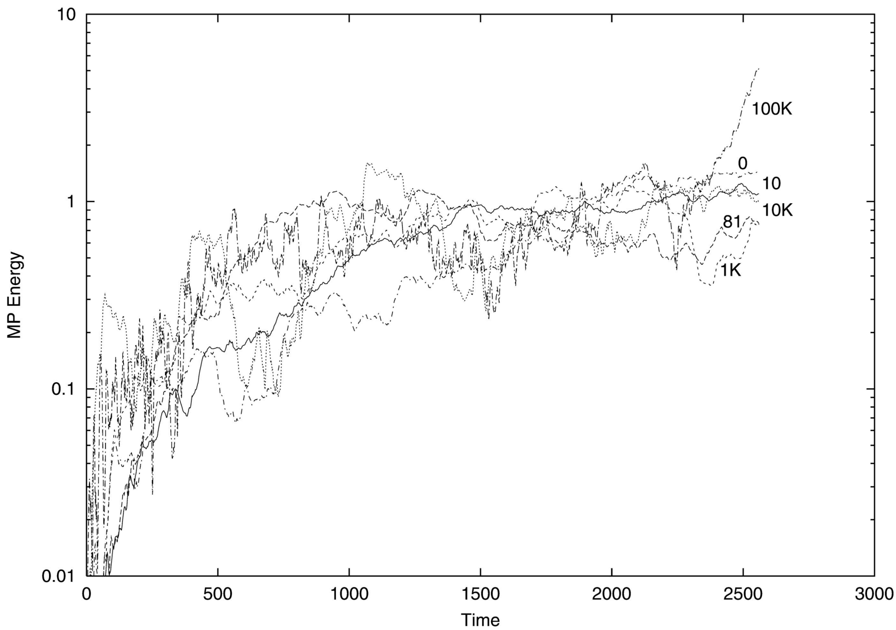



**Fig. 3.** Energies Weighted by Effective MP Masses

processes inherent in an equipartition picture. The scatter from experiment to experiment is about the same as the scatter within a given experiment in this plot. Call the "actual mass" $M_{\text{act}}$ and the "effective" mass $M_{\text{eff}}$.

## Compare Effective and Actual Masses

When the two kinds of "mass" are compared in Fig. 4, an interesting pattern emerges. The heavy diagonal line in that plot indicates equality of the effective mass and the actual mass, $M_{\text{eff}} = M_{\text{act}}$. Filled black squares indicate experiments from this $N = 128$ sequence. The other points come from the other sequences. The lines connecting these points are there simply to guide the eye.

There is a bit of a cheat in the abscissa of Fig. 4: $M_{\text{act}} = 0$ has been entered as about 1.4, to prevent the plotted points from disappearing off to the left. Otherwise both scales are logarithmic.

For small $M_{\text{act}}$, $M_{\text{eff}}$ is nearly constant at about 600. Once $M_{\text{act}}$ exceeds that value of 600, the two are nearly equal: $M_{\text{eff}} \approx M_{\text{act}}$.

To zero order, this set of experiments fits a pattern, $M_{\text{eff}} \approx \max(600, M_{\text{act}})$. There may be a dip partway along the constant $M_{\text{eff}}$ portion, but that hangs on just one experiment. The constant portion reflects the near–overlap of tracks noted in Fig. 2. It is not clear what physical features set the effective mass at about 600 for low $M_{\text{act}}$.

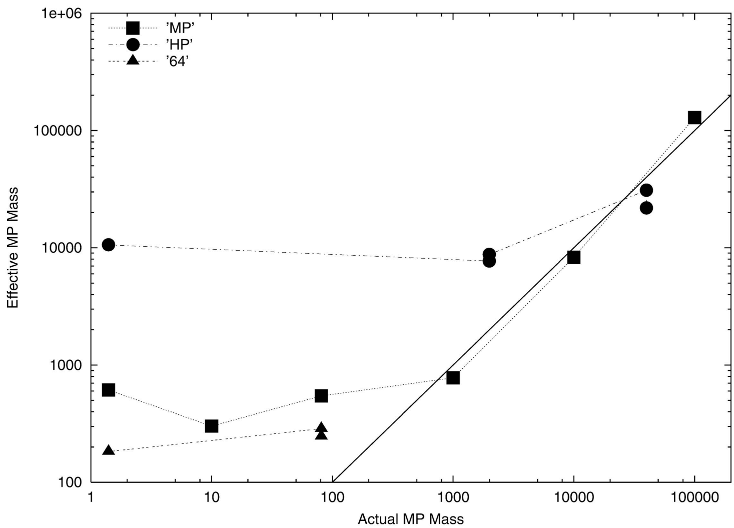



**Fig. 4.** Comparison of Effective MP Masses against Actual MP Masses

## 2.3 The Run with MP = 100K

The run with the very large $M_{\mathrm{act}} = 100\,000$ showed several features that set it apart from the other runs in the $N = 128$ sequence.

### Density Cusp Around the MP

Field particles near the MP form a definite density cusp. The cusp is shown as the solid track in Fig. 5. The dashed track comes from the run in which the MP mass is zero, for comparison. Both loads had the same number of particles and the same initial density contrast, $\rho_{\mathrm{max}}/\rho_{\mathrm{min}} = 100$, so the profile with 100K mass has much more mass at great distance.

This cusp is quite robust. It has a near power–law slope, $\rho \sim 17.5r^{-1.50}$, and the cusp remains substantially unchanged to the end of the experiment. However, it does not have much mass – only about 5000 field particles, some 5% of the mass of the MP.

The principal reason there is so little mass in the cusp is that so little phase volume is available. The cusp is confined to a small configuration volume, and the MP force field, which is softened with a softening length around 1.5 of our length units, restricts the velocities that can be bound. In analytic solutions (1), more mass accumulates around the MP with shorter softening in an isothermal configuration, finally diverging in the limit as the softening goes to zero.

The cusp seen in this run fits the expected density to fill the potential of the MP according to (1). There is no evidence of a nonlinear increase in cusp mass due to the mass in the cusp itself. Doubtless that increase must be present for more massive cusps.

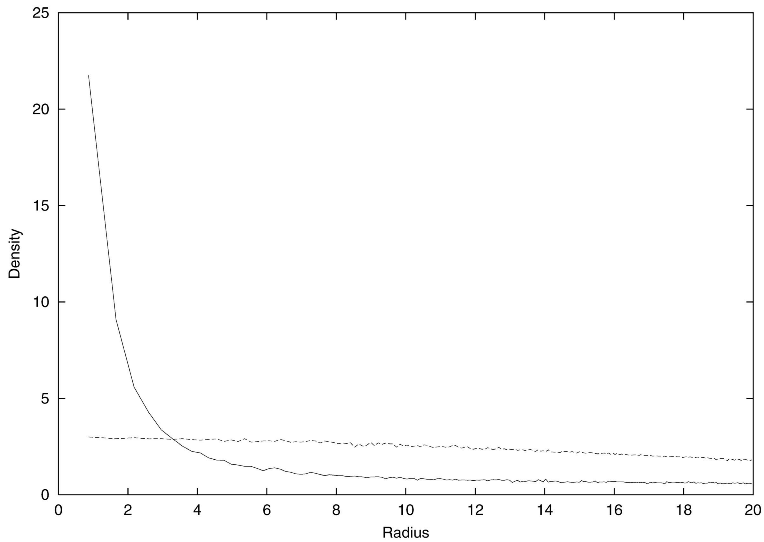



**Fig. 5.** Density Profiles at the Start of Experiments with MP mass = 100K (*solid line*), and with MP mass = 0 (*solid line*)

**The Cusp Moves With the MP.** A second property is that the cusp moves with the MP. It remains pretty well centered on the MP. This is shown in Fig. 6, in which the MP position and cusp position are plotted together at the same times. Figure 6 extends the time scale beyond that of Fig. 2. Data were filtered with a low–pass box filter because the raw data for the cusp position is noisy. Both the MP position and the cusp center positions were filtered by the *same* filter for this plot to avoid problems with filter calibration. Only the $x-$component of position is shown in Fig. 6, but agreement in the other two components is equally good. We regard the agreement as spectacular.

**Particles Bound to the Cusp.** A third property is that there seem to be NO particles permanently bound to the cusp. Passing field particles dwell there a bit longer, probably because their trajectories are curved, and thus account for what appears to be an enhanced density. This property can be seen from (1) if a density well is placed in an otherwise uniform isothermal sea of particles.

### The Large MP Orbit

Energy plots were cut off $T = 2560$ to avoid a large drift of the MP toward the end of the run. This drift turned out to be useful, even though it is harmless, because it helped to nail down the case that the cusp moves with the MP (Fig. 6).

### Features of MP Motion in the 100K Run

MP motions looked nearly oscillatory in runs with smaller MP masses, but they look quite irregular (albeit smooth) in the 100K run. There is noise in the MP

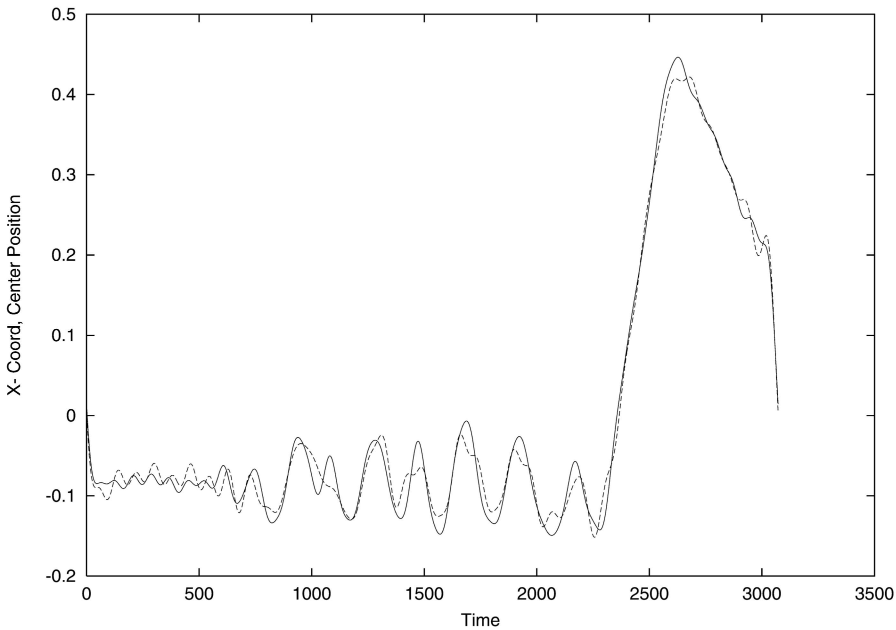



**Fig. 6.** Comparison of Cusp and MP Positions. (*solid line*) MP position, (*dotted line*) Cusp position

energy plots and in the MP velocity plots, but this is related to roundoff because of the small MP velocities in this run. Velocities are small because of the large mass.

The similarity of displacement and velocity plots broke down for the 100K run when the MP took off on its large drift, so the rule for estimating MP energy was no longer valid once that drift started. It was pretty good up to that time.

### 2.4 External Harmonic Potential

MP motions appear nearly harmonic, at least in the low-mass portions of Fig. 4, leading to the question how a batch of self–gravitating particles would act if placed within a fixed, fairly strong, harmonic potential.

This set of experiments was undertaken to investigate that question. The quick answer, as indicated by the filled black circle points in Fig. 4, is that systems in a harmonic background potential act much like systems without the harmonic potential, save that the effective mass along the low–mass portion is significantly higher at around 8 000. The double point at $M_{\rm act} = 2000$ represents two experiments. Points at $M_{\rm act} = 40\,000$ lie near the $M_{\rm eff} = M_{\rm act}$ line, just as the heavy points do in other experiments.

Tracks of the MP energy weighted by $M_{\rm eff}$ *vs.* time for this series of experiments are quite similar to those of Fig. 3, except that they are not as noisy. This is especially true of tracks with large $M_{\rm act}$.

An unusual aspect of the experiments in a strong external harmonic potential is that they all show a strong oscillation in the total kinetic energy (sum of all field particles plus the MP). The fractional amplitude of these oscillations ((peak to peak difference)/mean) was 9 to 12%, which is several times as great as the worst case encountered with other kinds of external potential. Fractional total KE oscillations were typically under 1% for other experiments, often mere fractions of a percent. Oscillations in the harmonic potential continue undamped for the duration of any experiment. The period of these oscillations was quite accurately half the period of the external harmonic potential, just the expected value. The design period of the external potential is set at 77 of our time units in all these experiments.

Lagrangian radii show the oscillation to have an amplitude proportional to the mean radius for that mass, a pattern that we have called the "fundamental" or "breathing" mode a different study [9]. The phase of the oscillation is the same at all masses.

These oscillations defied all efforts to get rid of them, through many trials and many checks. This suggests that there is something peculiar to harmonic potentials that makes it difficult to attain a Vlasov equilibrium. Of course, *everything* goes at the same frequency in a harmonic potential, so there are many resonances.

The period of MP oscillation in each of these experiments is around 74, measurably less than the period of 77 set by the external potential. Difference periods in the two kinds of oscillation are around 2000 of our time units. Our integrations extend beyond a period of the difference frequency, so MP motions are demonstably *not* in resonance with the frequency of the background potential.

The shorter period arises from the fact that the collection of field particles contributes a potential that itself looks harmonic, and which modifies the total frequency down near the center of the configuration. It increases the frequency, which decreases the period. Again, this is reasonable and expected.

## 2.5 Other Features

Some other results applied across the entire set of experiments reported here.

### Trajectory Separation

One feature was noticed in every one of the experiments of the three sets described here. As noted earlier, the MP was started from rest at the center of a "periodic cube." One field particle among the full load also started from rest at the same location. One of the routine post–run analyses consisted in tabulating the difference between the location of that one field particle and the MP throughout each experiment. The remarkable feature is that the field particle and the MP showed identical motions. Their position never deviated by more than roundoff in the accuracy with which the field particle position was tabulated: 0.002 of our length units at $N = 128$ and 0.001 at $N = 64$.

In a normal $n-$body problem, trajectories of these two particles would separate, and their distance (in phase space) would grow exponentially with time, like a Lyapunov exponent. We checked that this property holds for experiments run with grid codes like that used for the present experimental sequences, and found that exponential separation held in those cases as well. These tests were run some years ago.

These two trajectories do not separate. They remain very accurately together for the entire experiment. It is even more remarkable that they do not separate even in cases where $M_{\rm act} = 0$ – when, in effect, there is no MP.

This seems to indicate that motions near the MP are not chaotic, since otherwise the trajectories should separate. There seems to be a patch near the center that is locally harmonic, and harmonic potentials are not chaotic. With large values of $M_{\rm act}$ the softened potential of the MP is locally harmonic, and it produces a benign environment within which that particular field particle might move. In principle, that field particle might orbit around the MP with small amplitude, but if so, its amplitude is too small to be detected.

### MP Started with Nonzero Velocity

The MP started with nonzero velocity in experiments from another sequence. It settled down to the same amplitude of motion (same MP energy) as was attained by an MP starting from rest. This is consistent with any picture in which the steady state is described by equipartition. That other sequence is not described in this note. There was no external potential in that series of experiments.

## 3 Discussion

A few conclusions follow from the experimental results presented here.

### 3.1 The Relation Between Actual and Effective MP Masses

Experiments with low $M_{\rm act}$ show limited amplitude of MP motion which follows

$$M_{\rm eff} \approx \max(M_0, M_{\rm act}) \tag{3}$$

approximately, with differing $M_0$ depending on circumstances. We find $M_0 \approx 600$ for the $N = 128$ sequences, 250 for the $N = 64$ sequences, and 8 000 for the sequences with a strong harmonic external potential. So far, we have no way to estimate $M_0$, although trends seem reasonable.

The pattern at low $M_{\rm act}$ ($M_{\rm act} < M_0$) requires that the field particles and the MP be treated as a self–consistent whole. Smooth galaxy centers have a locally harmonic potential near the center, where a kind of collective effect takes over. This pattern confirms our earlier result [7, 8] since those experiments all had $M_{\rm act} = 0$. An interesting example was included in [10], where a group of 1000 particles, each with mass equal to that of a field particle, showed coherent motion that grew in a manner similar to that seen here. But it looks as if one early speculation is not valid: we guessed that a massive black hole might show displacements well in excess of those consistent with equipartition. It doesn't.

The break in (3) at high mass where $M_{\text{eff}} > M_{\text{act}}$ and the different character of MP motions with large $M_{\text{eff}}$ may be caused by a cusp, which could modify the force law in the neighborhood of the MP. The cusp moves with the MP, indicating that an effective mass for the combined MP and cusp should include the mass of the cusp. That enhancement is small in the experiments reported here, attaining a value at most around 5%, well below our ability to detect it.

An important aspect of these results is that the plateau at $M_0$ holds right down to $M_{\text{act}} = 0$ According to equipartition, the MP velocity should diverge in that case, but nothing of the sort happens.

**Remark.** The question how a batch of particles interact in a strong harmonic background potential might be illustrated by the following perturbation–type problem.

A batch of particles moves in a harmonic potential. They do not interact in the unperturbed state, but have a weak interaction is turned slowly. How do they respond?

The interesting case with $1/r$ Keplerian interaction is impossible to handle, so we seek an interaction with similar features that can be handled analytically. That leads to the following problem, which can be solved completely. Complete analytic solutions like this often suggest features which a more complex system must share.

Imagine a set of $n$ particles in a harmonic potential, but with a harmonic interaction between *absolutely every* pair of particles. Start with a one–dimensional problem. The formulation looks a lot like the normal modes problems discussed in mechanics texts.

Let the coordinate of the $i^{\text{th}}$ particle be $x_i$. All particles have the same mass and the interaction between the $i^{\text{th}}$ and $j^{\text{th}}$ particles is $\frac{1}{2}b(x_j - x_i)^2$. Characterize the background harmonic potential by its frequency, $\Omega$. Divide out the mass of each particle so the equation is per unit mass. The equation of motion for the $i^{\text{th}}$ particle (of a set of $N$ particles) is

$$\ddot{x}_i = -\Omega^2\, x_i + b \sum_{j \neq i}^{n} (x_j - x_i).$$

The sign on the interaction term puts the acceleration in the positive direction if $x_j > x_i$, which is what we want. The $j = i-$term has been omitted, but it can be included with no harm since it is always zero. Include it, and assume harmonic motion for all particles (with the explicit exponential, $e^{i\omega t}$ cancelled throughout).

Then the equation of motion (always per unit mass) can be written

$$(\omega^2 - \Omega^2 - nb)\, X_i + b \sum_{j=1}^{n} X_j = 0, \tag{4}$$

where $X_i$ and $X_j$ are the (complex) coefficients that multiply those $e^{i\omega t}$ terms. $\omega^2$ is an eigenvalue of the system.

In matrix notation,

$$MX = 0 \qquad \text{with} \qquad M = aI + bJ, \tag{5}$$

where $a = \omega^2 - \Omega^2 - nb$ and $b$ is dimensionally a frequency squared. The matrix $I$ is an $n \times n$ identity matrix and the matrix $J$ is the $n \times n$ matrix all of whose entries are 1's.

The matrix $J$ is the dyad, $f\,f^T$, of a $1 \times n$ column vector all of whose entries are 1's, so its rank is one. Superscript $T$ denotes the transpose of a matrix. $f$ is an eigenvector of $J$ with eigenvalue $n$. The eigenvalues of $J$ are all zero save for one whose value is $n$. It is straightforward to construct a complete basis for $J$, but it is not needed for present purposes.

Since $M = aI + bJ$ is real symmetric by construction, it can be reduced to diagonal form with real diagonal elements. There are $n-1$ matrix diagonal elements of $a$ and one whose value is $a + bn$. These statements are more or less evident, but the argument can be facilitated by the following considerations. Let $R$ be the ($n \times n$ orthogonal) matrix that diagonalizes $J$. Then

$$R^T\,M\,R = a\,(R^T\,I\,R) \;+\; b\,(R^T\,J\,R).$$

But $R^T\,J\,R$ is diagonal, and $R^T\,I\,R$ remains diagonal (since the identity remains an identity under any rotation), so this transformation also diagonalizes $M$. The diagonal elements add, term–by–term, so $M$ has $n-1$ diagonal elements of $a$ and one of $a + bn$.

This solves our problem, since (5) requires each diagonal element to be zero. When the defining values are plugged in, we have one eigenvalue with $\omega^2 = \Omega^2$ and $n-1$ eigenvalues of $\omega^2 = \Omega^2 + n\,b$.

The single eigenvalue, $\omega = \pm\Omega$ goes with the eigenvector $f$. All particles are in the same place and move together, oscillating with the frequency of the background harmonic potential. This solution, with a single large blob, should have been expected.

The set of $n-1$ degenerate eigenvalues goes at a higher frequency. At a given interaction strength their frequency grows without bound as $n$ becomes very large. Particles move in any of a variety of manners, save that they do not move as one large blob. “Modes” in the degenerate set feel the full acceleration due to all the particles, irrespective of where those particles might be.

The linear analysis singles out the property that all the particles, piled up at a single location and moving together, oscillate with the frequency of the background potential and don't feel the interactions. Any other motions feel the combined interaction of all other particles. This property may well be characteristic of this problem, but it might break down when the interactions become singular at zero separation.

An unpublished 1962 note by Joel L. Brenner [11], of Stanford Research Institute, pointed out how simple it is to manipulate matrices of the class encountered, $aI + bJ$. He also pointed out that the argument generalizes to block matrices. Sadly, his note is no longer available.

Because the harmonic oscillator problem is separable, the three dimensional extension is easy to construct. It gives rise to a block diagonal matrix, with three blocks each $n \times n$. The same arguments apply to each block. They would even apply for an anisotropic external oscillator potential if the problem is discussed in a coordinate system in which the potential is diagonal. Anisotropic $b$'s would have to be diagonal in the same frame, which is a bit artificial.

Unfortunately, the approach described here does not generalize easily to particles with different masses.

### 3.2 Theoretical Approaches

The problem of a massive particle in a sea of other particles, all embedded within a strong harmonic potential, looks a lot like the problem of Brownian motion in a harmonic potential. This problem has been discussed in the literature, apparently first by Ornstein. It is best known to astronomers through a 1943 paper by Chandrasekhar [12] but it was also included in [13]. That solution is one–dimensional, but a linearized three–dimensional extension appears in [14]. All three [12–14] are reprinted in [15]. The argument was recently retraced by Chatterjee [16]. All these approaches show a distribution of velocities and positions for the Brownian particle as a function of time. In a steady state, after transients have died out, expectation values are consistent with equipartition.

A Langevin equation is written in these treatments, with force on the Brownian particle treated as a series of impulses governed by a probability distribution. Those impulses are caused by close encounters, and field particles exert no forces on each other or on the Brownian particle other than the impulses. This picture is not completely applicable to the stellar dynamical problem because the underlying harmonic potential is generated by the field particles themselves, and the field particles interact, making the formulation of the theoretical problem somewhat inconsistent in the manner for self–gravitating particles.

Chatterjee *et al.* [16] included an $n-$body calculation and reported agreement with the Brownian motion picture. Because their field particles could not feel each other, they did not move self–consistently, so these authors did not find the plateau at low MP mass.

Calculations for this program were carried out at the NASA-Ames Research Center's NAS Systems Division under a grant of computing resources, which is gratefully acknowledged. We thank Dr. Bruce F. Smith of NASA-Ames Research Center for arranging the access. It is a pleasure to acknowledge helpful discussions with Prof. Kevin Prendergast.

## References

1. John Kormendy and Douglas Richstone. *Ann.Rev.Astronomy & Astrophysics*, **33**, 581–620, (1995)
2. L. Ferrarese and D. Merritt. *Phys. World*, **15, No. 6**, 41–46, (2002)
19. J. Binney and S. Tremaine. *Galactic Dynamics.* (Princeton University Press, Princeton, NJ., 1987)

4. S. Chandrasekhar. *An Introduction to the Study of Stellar Structure*. (Dover, New York, NY, 1957)
5. James M. Huntley and W. C. Saslaw. *Astrophys.J*, **199**, 328–335, (1975)
6. W. C. Saslaw. *Gravitational Physics of Stellar and Galactic Systems*. (Cambridge University Press, New York, 1985)
7. R. H. Miller and B. F. Smith. 'Goings-on at the center of a galaxy'. In D. J. Benney, F. H. Shu, and Chi Yuan, editors, *Applied Mathematics, Fluid Mechanics, Astrophysics, a Symposium to Honor C. C. Lin*, (World Scientific, Singapore 1988) pp. 366–372.
8. R. H. Miller and B. F. Smith. *Astrophys.J.*, **393**, 508–515, (1992)
9. R. H. Miller and B. F. Smith. *Celest. Mech. & Dyn. Astron*, **59**, 161–199, (1994)
10. R. H. Miller, G. R. Roelofs, and B. F. Smith. 'An experimental study of counter-rotating cores in elliptical galaxies'. In J. W. Sulentic, W. C. Keel, and C. M. Telesco, editors, *Paired and Interacting Galaxies, Proceedings of IAU Colloquium 124*, NASA Conference Publication 3098, (NASA GPO Washington DC, 1992) pp. 549–553.
11. J. L. Brenner. 'A set of matrices for testing computer programs'. Stanford Research Institute, Menlo Park, CA., 1962.
12. S. Chandrasekhar. *Reviews of Modern Physics*, **15**, 1–89, (1943)
13. G. E Uhlenbeck and L. S. Ornstein. *Phys.Rev.*, **36**, 823, (1930)
14. M. C. Wang and G. E. Uhlenbeck. *Reviews of Modern Physics*, **17**, 323–342, (1945)
15. Nelson Wax, editor. *Selected Papers on Noise and Stochastic Processes*, (Dover Publications, Inc. New York, 1954)
16. P. Chatterjee, Lars Hernquist, and A. Loeb. *Astrophys.J.*, **572**, 371–381, (2002)

# Weak Homology of Bright Elliptical Galaxies

Giuseppe Bertin

Università degli Studi, Dipartimento di Fisica, via Celoria 16, I-20133 Milano, Italy

**Abstract.** Studies of the Fundamental Plane of early-type galaxies, from small to intermediate redshifts, are often carried out under the guiding principle that the Fundamental Plane reflects the existence of an underlying mass-luminosity relation for such galaxies, in a scenario where elliptical galaxies are homologous systems in dynamical equilibrium. Here I will re-examine the issue of whether empirical evidence supports the view that significant systematic deviations from strict homology occur in the structure and dynamics of bright elliptical galaxies. In addition, I will discuss possible mechanisms of dynamical evolution for these systems, in the light of some classical thermodynamical arguments and of recent N-body simulations for stellar systems under the influence of weak collisionality.

## 1 Introduction

This article focuses on three main questions: (1) Are elliptical galaxies structurally similar to each other? (2) Which detailed dynamical mechanisms can make elliptical galaxies evolve? (3) Are there general trends to be anticipated for the evolution of these stellar systems?

Here I will report on a long-term research project aimed at providing answers to the above questions. Some interesting clues have been discovered only very recently [5], [6], [12]. Most of the paper refers to the class of bright ellipticals only; low-luminosity ellipticals are known to be characterized by different dynamical properties.

## 2 Structure of Bright Elliptical Galaxies

The answer to whether elliptical galaxies can be considered to be structurally similar to each other depends on the specific context in which the question is posed and addressed. Below, I will focus on the context of the physical interpretation of the Fundamental Plane ([26], [23]).

As demonstrated by a number of investigations (e.g., see [29], [30] for a study based on a sample of more than 200 early-type galaxies), the observed correlation that defines the Fundamental Plane, $\log R_e = \alpha \log \sigma_0 + \beta SB_e + \gamma$ (with $\alpha = 1.25 \pm 0.1$, $\beta = 0.32 \pm 0.03$, $\gamma = -8.895$ in the B band; the effective radius being measured in $kpc$, the central velocity dispersion in $km/sec$, the mean surface brightness in $mag/arcsec^2$ [2], [29]), is remarkably tight, with a scatter on the order of 15% in $R_e$.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 185–197, 2003.

The following simple physical argument has been put forward as an interpretation of this important physical scaling law. If we note that (1) the observed luminosity law of bright elliptical galaxies appears to be universal (the so-called $R^{1/4}$ law; [22]) and (2) the kinematical structure of these systems is regular and uniform ([4], [28]), it is natural to conclude that elliptical galaxies should be considered as homologous dynamical systems, in the sense that the relevant virial coefficient $K_V$ should be taken to be approximately constant from galaxy to galaxy. Then, (3) given the existence of the virial constraint, the Fundamental Plane can be seen as the manifestation of a mass–luminosity relation for galaxies (see [27], [46]). In fact, the virial theorem can be written as $GM_\star/R_e = L(G/R_e)(M_\star/L) = K_V\sigma_0^2$, where $M_\star$ is the mass of the luminous component and $L$ is the total luminosity. By eliminating $\sigma_0$ from the Fundamental Plane relation, one finds:

$$\left(\frac{M_\star}{L}\right)\frac{1}{K_V} \propto R_e^{(2-10\beta+\alpha)/\alpha}L^{(5\beta-\alpha)/\alpha} \sim L^{(5\beta-\alpha)/\alpha}. \tag{1}$$

The latter relation follows from the *empirical* fact that $2-10\beta+\alpha \approx 0$.

Unfortunately, there are empirical and theoretical findings that work against the hypotheses at the basis of the previous argument. First of all, significant deviations from the $R^{1/4}$ law have long been noted (see [17], [40]), and found to correlate systematically with the galaxy luminosity (see also [25]). Second, studies that have measured the amount and distribution of dark matter in ellipticals (see [4]) have shown that the presence of dark matter is more prominent in brighter and spatially larger galaxies, thus demonstrating that the virial coefficient may vary significantly from galaxy to galaxy. A curious theoretical point adds further caution to the perception that ellipticals should be considered homologous systems. This derives from direct inspection of the so-called $f_\infty$ sequence of models [9]. As demonstrated in [5], models that appear to be all (see Fig. 1, for $\Psi$ in the range $7-10$) very well fitted by the $R^{1/4}$ law, over a luminosity range of more than ten magnitudes, may be characterized by significantly different values of the relevant virial coefficient (see Fig. 2, the triangles representing the virial coefficient for the $f_\infty$ sequence of models), as a result of the impact of a more and more concentrated nucleus.

In [5] we have further confirmed, by close inspection of four cases (NGC 1379, NGC 4458, NGC 4374, NGC 4552; studied in great detail by comparing the performance of a number of fitting procedures on data taken from [19], [18]), that the Sersic [41] index $n$ for the $R^{1/n}$ photometric profiles can indeed be very different from 4 (in particular, for NGC 4552 we find $n \approx 11$, with residuals on the order of 0.2 magnitudes; a fit performed with the $R^{1/4}$ law would lead to residuals up to one magnitude, while a fit based on an $R^{1/4}+$ exponential profile would have residuals up to half a magnitude). On the other hand, we have checked that, if the luminosity range where the fit to the photometric profile is performed is reduced to less than 5 magnitudes, then (see [16]) the profiles are indeed well fitted by a "universal" $R^{1/4}$ law.

In conclusion, while we find it necessary to dismiss strict homology as a viable description of elliptical galaxies in relation to the interpretation of the

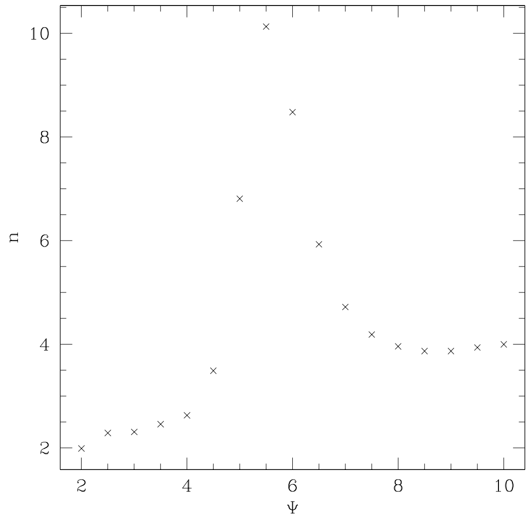



**Fig. 1.** The best-fit $n(\Psi)$, obtained by fitting the $f_\infty$ models, projected along the line of sight, with $R^{1/n}$ profiles. Note the plateau at $n = 4$ reached by concentrated (high-$\Psi$) models, for which the radial range adopted in the fit is $0.1 \leq R/R_e \leq 10$ (from [5])

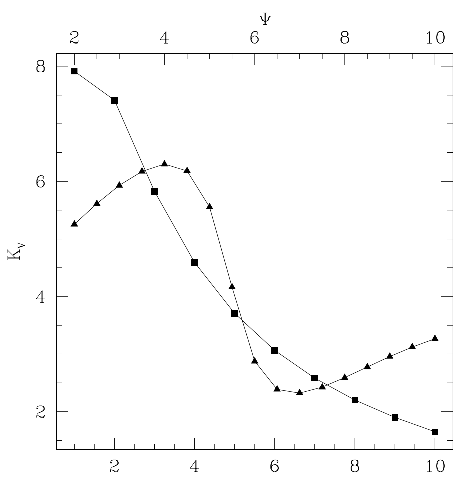



**Fig. 2.** The virial coefficient for the $f_\infty$ (triangles) and for the isotropic $R^{1/n}$ (squares) models, based on an aperture of radius $R_e/8$ (from [5])

Fundamental Plane, the existence of the empirical scaling law suggests that some kind of *weak homology* must be enforced (expressed by (1)), as a correlation between structural properties and total luminosity. In [5] we have also proved that a large scatter in the dynamical correlations (e.g., in the $n \sim -19 + 3 \log L$

relation noted in [17], [25]) may well be compatible with the observed tightness of the Fundamental Plane.

## 3 Mechanisms of Dynamical Evolution

Given the conclusion that elliptical galaxies have to be considered only weakly homologous systems, it is natural to ask whether and how individual galaxies may change their internal structure via dynamical processes. This general issue is especially important, if we recall that typically, in the study of the cosmological evolution of the Fundamental Plane (see [44] and references therein), strict homology and thus a mass–luminosity relation is assumed for the observed galaxies and an interpretation of the data (see Fig. 3) is made in terms of pure *passive evolution* (through the evolution of the luminosity resulting from the evolution of the properties of stellar populations).

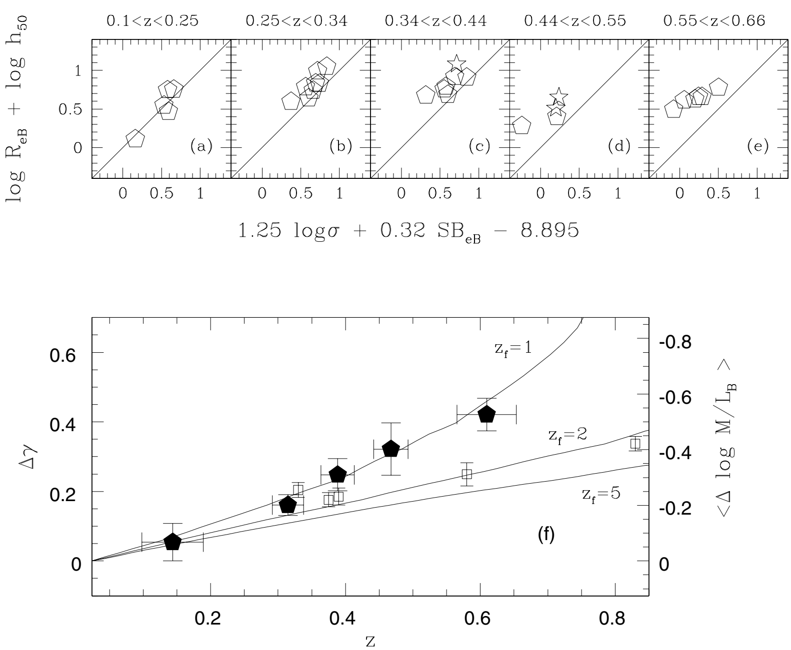



**Fig. 3.** The Fundamental Plane in the rest frame B band. In panels (a) to (e), field E/S0 galaxies are shown, binned in redshift, and compared to the Fundamental Plane found in the Coma Cluster by [2]. Panel (f) shows the average offset of the intercept of field galaxies from the local Fundamental Plane relation as a function of redshift (large filled pentagons) compared to the offset observed in clusters (open squares). The solid lines represent the evolution predicted for passively evolving stellar populations formed in a single burst at $z = 1, 2, 5$ (from top to bottom). This figure is taken from [44] where full references are given to the sources for cluster data points and stellar synthesis models

Besides the possibility of major merger events, are there significant sources of dynamical evolution for elliptical galaxies to be considered? As noted recently [3], the traditional approach to the study of elliptical galaxies, in terms of equilibrium and stability for the solutions of the collisionless Boltzmann equation, supplemented by the Poisson equation, may be misinterpreted. Given the very large values of typical star-star relaxation times in elliptical galaxies (see [20], [42]) it is generally taken for granted that, unless a system happens to be in a dynamically unstable state (for example, a condition of excessive radial anisotropy; see [39]), its state is basically "frozen" into an equilibrium distribution function. Thus the only task left to the dynamicist would be to decipher which distribution function best describes the observed states (a task that is particularly difficult for non-spherical systems) taken to be strictly stationary.

In our opinion, the above picture is oversimplified and may lead to an improper perception of the dynamics of real stellar systems. If, for simplicity, we take the view that elliptical galaxies have formed via collisionless collapse (see [45]), we should realize that splitting past and present conditions (that is formation processes and a collisionless equilibrium state) is just an idealization that the theory makes in order to define a basic state and to study its properties. In reality, stellar systems evolve continually and we should check to what extent the evolution processes change the internal structure of galaxies.

There are several specific mechanisms and causes for dynamical evolution that could be studied: (i) "Granularity" in phase space left over from the initial collapse; (ii) Presence of gas in various phases, especially of the hot X-ray emitting interstellar medium; (iii) Interactions with a compact central object; (iv) Interactions between the galaxy and its own globular cluster system; (v) Interactions with external satellites and effects of tides and minor mergers.

In a recent paper [6] we have tried to quantify the role of items (iv) and (v) above by means of N-body simulations. The idea at the basis of these studies is that heavy objects can suffer dynamical friction and then be dragged in toward the galaxy center, as studied earlier, for example, in [15], [14], [48]; in fact, the parallel momentum transport relaxation time $T_{fr}$ is related to the deflection relaxation time $T_D$ by a factor that can be very small when a heavy test particle moves through a field of lighter particles: $T_{fr} = 2T_D m_f/(m_t + m_f)$. We have thus revisited the problem of simulating the orbital decay of a satellite, placed initially on a circular orbit at the periphery of a galaxy, and basically confirmed the general findings presented in [14]; note that our simulations have been made with about one million particles, while the earlier simulations had been carried out with a few thousand particles. Then we have proceeded to address a quasi-spherical problem in which the satellite is fragmented into many smaller objects (several runs were made with either 20 or 100 fragments), distributed on a spherical shell. The quasi-spherical symmetry that characterizes this study has the important advantage of allowing for a smoother framework, basically free from other effects unrelated to dynamical friction, such as those associated with lack of equilibrium in the initial configuration. Furthermore, with respect to the earlier studies of the orbital decay of a single satellite, our attention here

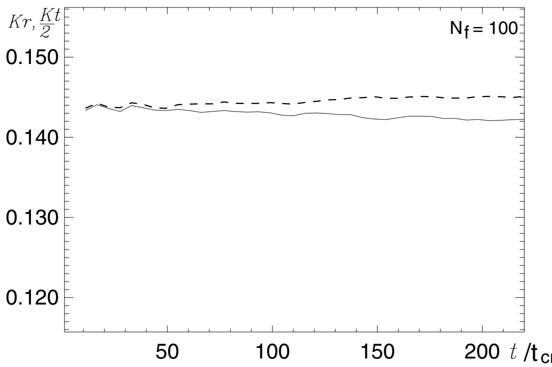



**Fig. 4.** The development of pressure anisotropy in a galaxy as a result of the interaction with a shell of $N = 100$ fragments dragged in toward the galaxy center by dynamical friction. The broken line represents the evolving value of $K_T/2$, where $K_T$ is the total kinetic energy associated with the star motions in the tangential directions; the solid line represents the evolving value of $K_r$, the total kinetic energy associated with the star motions in the radial direction (from [6])

is mostly shifted to measuring the evolution of the underlying structure of the hosting galaxies. One effect observed, while the fragments are slowly dragged in toward the center, is a general change in the stellar density distribution with respect to the initial polytropic basic state. Another expected effect that we have been able to quantify, starting from an initially isotropic distribution of stellar orbits, is the slow growth of a tangentially biased pressure anisotropy (see Fig. 4). All these slow dynamical evolution effects appear to be genuinely associated with the process of dynamical friction exerted by the stars on the minority component of heavier objects. We are planning a survey of cases that should allow us to identify general properties of dynamical evolution in elliptical galaxies resulting from the interaction between the stars and a significant population of globular clusters or of the merging of a large number of small satellites.

## 4 General Trends from Thermodynamical Arguments

In order to study possible general trends that may be anticipated for the evolution of elliptical galaxies, we refer to the general framework that has been successfully applied to the context of the evolution of globular clusters. Globular clusters appear to be well represented by King [34] models (see [24]). They are recognized to be non-homologous stellar systems, subject to dynamical evolution resulting from internal effects (such as weak collisionality and evaporation) and external perturbations (such as disk-shocking, when, in our Galaxy, their orbits happen to cross the disk). It has been noted that these mechanisms of dynamical evolution make a globular cluster evolve approximately along the King equilibrium sequence (see [47] and references therein).

For globular clusters, an important paradigm is provided by the *gravothermal catastrophe* [37], which offers interesting applications and physical interpretation (for a review, see [42]). Here we recall that, starting from the study of isothermal gas spheres [13], the gravothermal catastrophe is expected to occur also in stellar systems (see [1], [37]). The instability is interpreted as due to the curious property of self-gravitating systems of being characterized by a negative effective specific heat. Although for stellar systems a rigorous proof has been provided only for idealized models where an isothermal set of stars is confined by a spherical box, the paradigm is generally believed to be sufficiently robust to be applicable to real stellar systems, provided that they possess a sufficient level of internal collisionality. An independent element that strengthens the view that the paradigm is indeed robust has been added by an analysis that has shown, for an isothermal gas, that spherical symmetry is not a necessary ingredient [35].

Following some arguments initially put forward by Lynden-Bell (see [36], [37]), would there be a way to lay out a similar scenario for elliptical galaxies as partially relaxed stellar systems? If so, we would gather powerful "thermodynamical" arguments to determine general trends for evolution, beyond the specific paths produced by a given dynamical mechanism.

In our view, there are two aspects of the problem that require clarification. A first point is that we would like to start from a physically justified equilibrium sequence, much like King models for globular clusters, able to describe the general properties of elliptical galaxies. A second point is that, formally, the origin of the gravothermal catastrophe can be traced to the Poincaré stability of linear series of equilibria (see [31], [32]). For a proper mathematical derivation, one would thus like to start from a sequence of collisionless models derived rigorously from the Boltzmann entropy. In the absence of such a sequence, a derivation of the gravothermal catastrophe has been based on either an *unjustified ansatz* (see [33], [38]), that the global temperature of the system would be associated with the coefficient multiplying the energy in the distribution function, or the use of non-standard entropies [21] (but for unrealistic models).

In order to address the first point, we may refer to a sequence of models that have been found to be very promising for a realistic description of elliptical galaxies (the so-called $f_\infty$ models; [9], see the review [11]). These models have been inspired by the characteristics of the products of collisionless collapse, as derived from N-body simulations [45]. In the simple spherical case, they are based on the distribution function $f_\infty = A(-E)^{3/2}\exp(-aE - cJ^2/2)$, with $A, a, c$ positive constants, and define a one-parameter equilibrium sequence, which, much like King models, can be parameterized in terms of the dimensionless central potential $\Psi = -a\Phi(0)$. For positive values of $E$ the distribution function is taken to vanish. When $\Psi$ increases beyond a certain value, around $\Psi = 7$, the models have a projected mass density profile that is well fitted by the $R^{1/4}$ law and indeed they turn out to be an excellent tool to fit the observations. From the point of view of statistical mechanics, they have been found [43] to be compatible with a derivation based on a partition of phase space in terms of the star energy and the star angular momentum square, under the assumption that detailed conservation of the star angular momentum is required at large values of angular momentum. This closely follows our understanding of the process of partial violent relaxation [36]. Unfortunately, the derivation is based on heuristic arguments and the distribution function does not follow from a straightforward exact mathematical extremization of the Boltzmann entropy; in particular, the orbit time that acts as a weight to the cells in phase space is replaced, for simplicity, by a factor $1/(-E)^{3/2}$, which is approximately correct only for weakly bound orbits. Therefore, attempts at using this equilibrium sequence to study the gravothermal catastrophe in the context of elliptical galaxies, while definitely appealing from the physical point of view (see also [8]), would remain less satisfactory from the formal point of view.

Now we have shown [12] that we can carry out a program that is satisfactory not only from the physical point of view (because it is based on an equilibrium sequence, also inspired by studies of collisionless collapse [45], that is able to match the properties of observed galaxies), but also from the mathematical point of view (because the distribution is derived rigorously from the Boltzmann entropy by requiring the conservation of a third global quantity $Q$, in addition to total energy and total mass). The program is made possible by the second option explored in [43] for the construction of models of partially relaxed stellar systems. This option leads to the so-called $f^{(\nu)}$ models. It was already noted [43] that the general physical properties of the $f^{(\nu)}$ models are close to those of the $f_\infty$ models and, in particular, that for $\nu$ in the range $0.5-1$ their projected mass distribution, for concentrated models, follows the $R^{1/4}$ law.

Let $f$ be the single-star distribution function, $E$ the single-star specific energy, and $J$ the magnitude of the single-star specific angular momentum. Consider the standard Boltzmann entropy:

$$S = -\int f \ln f d^3v d^3x \tag{2}$$

and extremize it under the constraint that the total mass

$$M = \int f d^3v d^3x, \tag{3}$$

the total energy

$$E_{tot} = \frac{1}{3}\int E f d^3v d^3x, \tag{4}$$

and a third global quantity

$$Q = \int J^\nu |E|^{-3\nu/4} f d^3v d^3x \tag{5}$$

are assigned. Then the resulting distribution function is

$$f^{(\nu)} = A \exp\left(-aE - dJ^\nu |E|^{-3\nu/4}\right). \tag{6}$$

In the above expression, the quantities $A$, $a$, and $d$ are positive constants. The parameter $\nu$ is a free (positive) parameter, which was argued [43] to be in the range $0.5 - 1.0$. In the following we refer to the case $\nu = 1$. Note that the three constants appearing in the distribution function define two scales and one dimensionless parameter, which we take to be $\gamma = ad^{2/\nu}/(4\pi GA)$.

Self-consistent models generated by such distribution function are computed from the Poisson equation, solved under the boundary conditions of regular potential at the center and of Keplerian potential at very large radii. For positive values of $E$ the distribution function is taken to vanish. If we introduce the dimensionless central potential $\Psi = -a\Phi(0)$, the outer boundary condition defines a sort of eigenvalue problem that is solved by the relation $\gamma = \gamma(\Psi)$. The self-consistent models thus make a one-parameter equilibrium sequence.

By careful numerical integration, one may then proceed to calculate the functions $S = S(M, Q, \Psi)$ and $E_{tot} = E_{tot}(M, Q, \Psi)$ on the equilibrium sequence (see Fig. 5) and from here the inverse *global temperature*

$$\zeta = \left(\frac{\partial S}{\partial E_{tot}}\right)_{M,Q} . \tag{7}$$

The onset of the gravothermal catastrophe is thus determined by inspection of the equilibrium sequence studied in the $(E_{tot}, \zeta)$ plane (following [31]).

In [12] we have implemented the above program and shown that for the $f^{(\nu)}$ models the gravothermal catastrophe is expected to set in at $\Psi \approx 9$. Surprisingly, around this value of the concentration, the projected mass distribution turns out

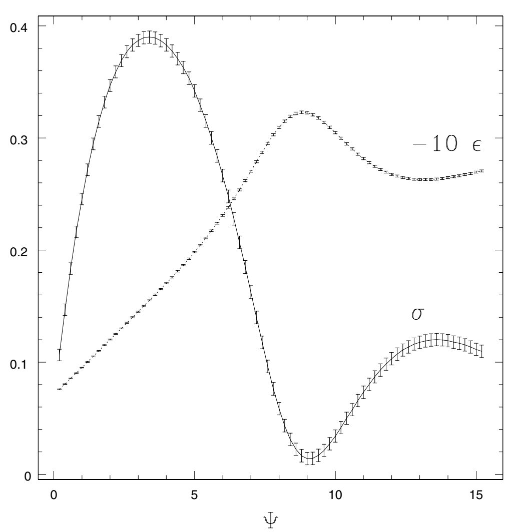



**Fig. 5.** Specific entropy and total energy along the equilibrium sequence of $f^{(\nu)}$ models with $\nu = 1$ (as a function of the concentration parameter $\Psi$, at constant $M$ and $Q$, and thus expressed by means of the dimensionless functions $\sigma(\Psi)$ and $\epsilon(\Psi)$). Note that for $\Psi < 3.5$ the models are characterized by a negative global temperature, because the derivatives of $S$ and $E_{tot}$ have opposite signs. This figure is taken from [12]

to be very well fitted by the $R^{1/4}$ law (this general point had already been noted in [43], but outside the context of the gravothermal catastrophe). For values of $\Psi$ close to and beyond 9, the general properties of the instability "spiral" in the $(E_{tot}, \zeta)$ plane, based on the proper thermodynamical definition of the global temperature, are the same as in the $(E_{tot}, \hat{a})$ plane, based on the *ansatz* that the temperature of the models is determined by the coefficient $a$ (see Fig. 6).

One important point noted in [12] is a qualitative departure of the behavior of the instability "spiral" at low values of $\Psi$. For the original gas sphere and for the stellar dynamical analogue of a stellar system confined by a box with reflecting walls, the limit of low concentration was identified as that of a *non-gravitating ideal gas*, subject to Boyle's law. In our case, the analogy breaks down. In fact, the global temperature turns out to *change sign* at $\Psi \approx 3.5$ (see Fig. 5). Such a drastic event should be accompanied by some physical counterpart in the dynamical behavior of the system. Surprisingly, the value of $\Psi \approx 3.5$ coincides with that for the threshold of the radial-orbit instability [39] (for the context of $f_\infty$ models, see [7] and [10]). In other words, by undertaking a thermodynamical description of the equilibrium sequence of models defined by the $f^{(\nu)}$ distribution function, we have found arguments that lead us naturally not only to the interpretation of the observed $R^{1/4}$ law, but also to one clue for the interpretation of the radial-orbit instability of collisionless stellar systems. Besides the proper-

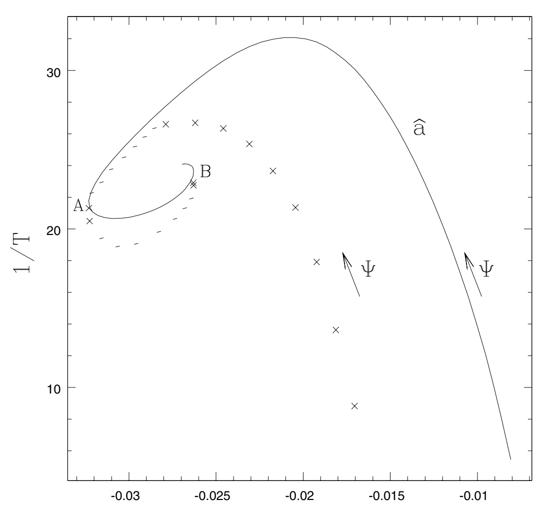



**Fig. 6.** The instability "spiral" of $f^{(\nu)}$ models with $\nu = 1$. The solid line refers to the results obtained from the *ansatz* that the coefficient $a$ represents the inverse global temperature. Crosses represent the inverse global temperature from the definition $\partial S/\partial E_{tot}$; other symbols indicate estimated points for which the adopted numerical differentiation is less reliable. Point A marks the onset of the gravothermal catastrophe (from [12])

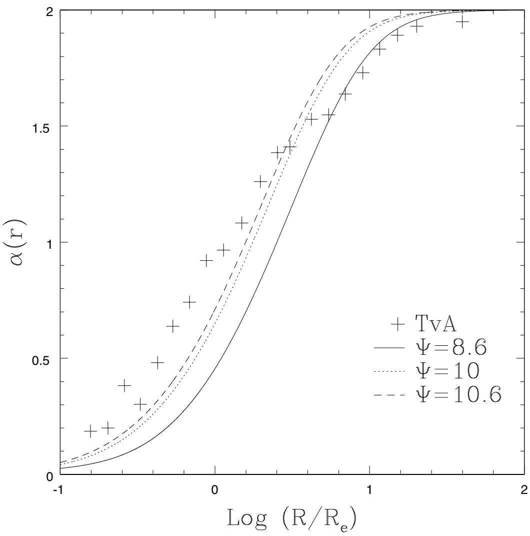



**Fig. 7.** Pressure anisotropy profiles $\alpha = 2 - (\langle v_\phi^2 \rangle + \langle v_\theta^2 \rangle)/\langle v_r^2 \rangle$ as a function of radius for selected $f^{(\nu)}$ models ($\nu = 1$) compared to the pressure anisotropy profile found [45] in numerical simulations of collisionless collapse. This figure has been prepared by M. Trenti

ties just outlined, one important additional aspect that makes the $f^{(\nu)}$ models, at this point, more appealing than the $f_\infty$ models is their anisotropy level. We had noted (e.g., see [11]) that the $f_\infty$ models are actually too isotropic, when compared with the final products of simulations of collisionless collapse [45]. The present models turn out to be much more interesting even in this respect. We have indeed checked that their characteristic anisotropy profile, for values of $\Psi$ close to the onset of the gravothermal catastrophe, is very similar to that observed in the numerical simulations (see Fig. 7).

## 5 Conclusions

For a physically justified family of equilibrium models, representing the result of incomplete violent relaxation, and derived rigorously from the Boltzmann entropy, we have shown that, at high concentration values, the onset of the gravothermal catastrophe is found to occur at $\Psi \approx 9$, in the parameter domain where models are characterized by an $R^{1/4}$ projected density distribution. At low concentration values, the equilibrium sequence presents a drastic departure from the limit of the classical isothermal sphere, because models become associated with a negative global temperature. The transition point, $\Psi \approx 3.5$, turns out to coincide with the point of the sequence where the radial-orbit instability sets in. In the intermediate concentration regime, $3.5 < \Psi < 9$, the structural properties of the models change, much like those of models along the

King equilibrium sequence, a family of models that is known to capture the non-homologous properties of globular clusters. It is our hope that, in this domain of intermediate concentration values, the $f^{(\nu)}$ models may be used to describe the characteristics of weak homology of elliptical galaxies.

**Acknowledgements**

I would like to thank L. Ciotti, M. Del Principe, T. Liseikina, M. Lombardi, F. Pegoraro, M. Stiavelli, M. Trenti, T. Treu, and T. van Albada, for their collaboration in this work, and the organizers of the workshop"Galaxies & Chaos", for their warm hospitality in Athens. This work has been partially supported by MIUR of Italy.

## References

1. V.A. Antonov: Vestnik Leningr. Univ. no. 19, 96 (1962) (Engl. Transl.: in *Structure and Dynamics of Elliptical Galaxies*, ed. by T. de Zeeuw (Reidel, Dordrecht 1986) pp. 531-548
2. R. Bender, R.P. Saglia, B. Ziegler et al.: Astrophys. J. **493**, 529 (1998)
3. G. Bertin: 'Gravitational plasmas'. In *Plasmas in the Universe, 142nd Course of the International School of Physics "Enrico Fermi"*, ed. by B. Coppi, A. Ferrari, E. Sindoni (Società Italiana di Fisica, Bologna 2000) pp. 373-393
4. G. Bertin, F. Bertola, L.M. Buson et al.: Astron. Astrophys. **292**, 381 (1994)
5. G. Bertin, L. Ciotti, M. Del Principe: Astron. Astrophys. **386**, 149 (2002)
6. G. Bertin, T. Liseikina, F. Pegoraro: paper submitted (2002)
7. G. Bertin, F. Pegoraro, F. Rubini, E. Vesperini: Astrophys. J. **434**, 94 (1994)
8. G. Bertin, R.P. Saglia, M. Siavelli: 'Spiraling into the $R^{1/4}$ law of ellipticals'. In *New ideas in Astronomy*, ed. by F. Bertola, J.W. Sulentic, B.F.Madore (Cambridge University Press, Cambridge 1988) pp. 93-95
9. G. Bertin, M. Stiavelli: Astron. Astrophys. **137**, 26 (1984)
10. G. Bertin, M. Stiavelli: Astrophys. J. **338**, 723 (1989)
11. G. Bertin, M. Stiavelli: Rep. Prog. Phys. **56**, 493 (1993)
12. G. Bertin, M. Trenti: Astrophys. J. **584**, in press (2003)
13. W.B. Bonnor: Mon. Not. Roy. Astron. Soc. **116**, 351 (1956)
14. Tj.R. Bontekoe: Orbital decay of satellite galaxies. PhD Thesis, Groningen University, Groningen (1988)
15. Tj.R. Bontekoe, T.S. van Albada: Mon. Not. Roy. Astron. Soc. **224**, 349 (1987)
16. A. Burkert: Astron. Astrophys. **278**, 23 (1993)
17. N. Caon, M. Capaccioli, M. D'Onofrio: Mon. Not. Roy. Astron. Soc. **265**, 1013 (1993)
18. N. Caon, M. Capaccioli, M. D'Onofrio: Astron. Astrophys. Suppl. **106**, 199 (1994)
19. N. Caon, M. Capaccioli, R. Rampazzo: Astron. Astrophys. Suppl. **86**, 429 (1990)
20. S. Chandrasekhar: Astrophys. J. **97**, 251 (1943)
21. P.H. Chavanis: Astron. Astrophys. **386**, 732 (2002)
22. G. De Vaucouleurs: Ann. d'Astrophys. **11**, 247 (1948)
23. S. Djorgovski, M. Davis: Astrophys. J. **313**, 59 (1987)
24. S. Djorgovski, G. Meylan: Astron. J. **108**, 1292 (1994)

25. M. D'Onofrio, M. Capaccioli, N. Caon: Mon. Not. Roy. Astron. Soc. **271**, 523 (1994)
26. A. Dressler, D. Lynden-Bell, D. Burstein et al.: Astrophys. J. **313**, 42 (1987)
27. S.M. Faber, A. Dressler, R.L. Davies et al.: 'Global scaling relations for elliptical galaxies and implications for formation'. In *Nearly normal galaxies: From the Planck time to the present*, ed. by S.M. Faber (Springer, New York 1987) pp. 175-183
28. O. Gerhard, A. Kronawitter, R.P. Saglia, R. Bender: Astron. J. **121**, 1936 (2001)
29. I. Jørgensen, M. Franx, P. Kjærgaard: Astrophys. J. **411**, 34 (1993)
30. I. Jørgensen, M. Franx, P. Kjærgaard: Mon. Not. Roy. Astron. Soc. **280**, 167 (1996)
31. J. Katz: Mon. Not. Roy. Astron. Soc. **183**, 765 (1978)
32. J. Katz: Mon. Not. Roy. Astron. Soc. **189**, 817 (1979)
33. J. Katz: Mon. Not. Roy. Astron. Soc. **190**, 497 (1980)
34. I.R. King: Astron. J. **71**, 64 (1966)
35. M. Lombardi, G. Bertin: Astron. Astrophys. **375**, 1091 (2001)
36. D. Lynden-Bell: Mon. Not. Roy. Astron. Soc. **136**, 101 (1967)
37. D. Lynden-Bell, R. Wood: Mon. Not. Roy. Astron. Soc. **138**, 495 (1968)
38. M. Magliocchetti, G. Pucacco, E. Vesperini: Mon. Not. Roy. Astron. Soc. **301**, 25 (1998)
39. V.L. Polyachenko, I.G. Shukhman: Sov. Astron. **25**, 533 (1981)
40. P. Prugniel, F. Simien: Astron. Astrophys. **321**, 111 (1997)
41. J.L. Sersic: *Atlas de galaxias australes* (Observatorio Astronomico, Cordoba 1968)
42. L. Spitzer: *Dynamical evolution of globular clusters* (Princeton University Press, Princeton 1987)
43. M. Stiavelli, G. Bertin: Mon. Not. Roy. Astron. Soc. **229**, 61 (1987)
44. T. Treu, M. Stiavelli, S. Casertano et al.: Astrophys. J. Lett. **564**, 13 (2002)
45. T.S. van Albada: Mon. Not. Roy. Astron. Soc. **201**, 939 (1982)
46. T.S. van Albada, G. Bertin, M. Stiavelli: Mon. Not. Roy. Astron. Soc. **276**, 1255 (1995)
47. E. Vesperini: Mon. Not. Roy. Astron. Soc. **287**, 915 (1997)
48. M.D. Weinberg: Mon. Not. Roy. Astron. Soc. **239**, 549 (1989)

# Observing Chaos in Disk Galaxies*

Preben Grosbøl

European Southern Observatory, Karl-Schwarzschild-Str. 2, 85748 Garching, Germany

**Abstract.** Regions in disk galaxies where one would expect to find chaotic behavior are likely to be associated with major stellar resonances such as the co-rotation. The possible identification of such locations in real galaxies is illustrated by examples of four spiral galaxies observed in the K band. Observational issues related to the detection of chaotic regions are discussed. Although surface photometry may suggest chaotic regions, it is essential to compare detailed velocity profiles with dynamic models to estimate the probability of such claims. Finally, the feasibility of performing observations of chaos with current state-of-the-art facilities such as the VLT is considered. It is found that it should be possible down to a surface brightness level of I $\approx$ 20 mag/arcsec$^2$ corresponding to the end of bars in typical disk galaxies whereas access to detailed studies of chaos in the main spiral pattern would require more powerful facilities.

## 1 Introduction

Chaotic behavior of orbits in galactic potentials is frequently seen in analytical models and numerical experiments (e.g. N-body simulations). By increasing perturbations in models of spiral galaxies, one can observe a transition of stable orbits to chaotic ones [4]. Numeric techniques make it simple to identify such orbits by calculating their dynamic spectra [29]. The existence of chaotic behavior in models does not automatically mean that it is an important phenomena in real galaxies. It is also possible that growing spiral modes are damped by non-linear effects, causing an increased velocity dispersion, before a significant fraction of the stars becomes chaotic. Thus, it is of high interest to estimate the level of chaotic behavior in disk galaxies.

The current paper considers the possibility of observing chaos in real galaxies. The next section looks on the importance of the environment while the main regions where one may expect chaotic behavior are identified in Sect. 4. Different indicators for chaos in disk galaxies, including our own, are discussed in the following sections. The feasibility of actually observing chaos is then considered in the last section using VLT instrumentation as a reference.

## 2 Environment

Current disk galaxies would have had time enough to reach a relative stable and relaxed state if they were formed in a single collapse at an early epoch of the

* Based on observations collected at the European Southern Observatory, La Silla, Chile.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 201–212, 2003.

universe. In a Cold Dark Matter scenario where galaxies are formed in hierarchical mergers over a longer time, it is less obvious that present time galaxies are relaxed, isolated systems as frequently assumed in models. Thus, it is of interest to verify whether typical nearby spiral galaxies have been able to achieve a quasi-stationary state or not.

The history of star formation in field galaxies provides some information on the environment at earlier epochs since a higher frequence of encounters between galaxies may yield an increased star formation rate. Madau et al. [21] used data from the Hubble Deep Field (HDF) to estimate the past star formation rate and found that it increased back to around a redshift of z≈1.5. At higher redshifts, the star formation rate is either close to constant or still monotonically increasing [19] depending on the exact corrections applied for dust attenuation and cosmological surface brightness dimming which are significant at these redshifts. Although this does not exclude a monolithic collapse scenario, it would predict a higher metal mass density at high redshifts than observed for absorber in quasi-stellar objects.

Another approach is to study the morphology of galaxies as function of their redshift as done by van den Berg [10] who analyzed the HDFs. He found that galaxies with z<1 appear largely as disk-like while those with z>2 either had chaotic knots or were centrally concentrated which may be precursors to elliptical galaxies or bulges in spirals. For the higher redshifts, a larger fraction of galaxies showed evidence of recent mergers. Considering the star formation rate at earlier epochs discussed above, van den Bergh [10] suggested that a change of slope around z≈1.5 may be cause in a transition from a merger driven star formation at early epochs to one mainly occurring in galactic disks.

The Hubble morphological classification scheme was found satisfactory for redshifts z<0.5 while it at higher redshift became increasingly difficult to apply as the fraction of peculiar galaxies got larger [10]. The frequency of barred spiral galaxies also became smaller at redshifts z>0.5 which could be caused by their disks being hotter and therefore more stable against bar instabilities. Spiral structure observed at these higher redshifts also appear more chaotic and less well developed.

The influence of the environment on the internal structure of disk galaxies was investigated by van den Bergh [8, 9] who looked on a set of 930 Northern galaxies in the Revised Shapley-Ames catalog [24]. From this sample, he found no statistical significant change in the morphology of bars or spiral structures as function of their environment which suggests that such structures primary depend on the general properties of the parent galaxy. The merger rate of local disk galaxies was estimated by Keel & Wu [16] to be 4.2 per Hubble time for pairs and 0.33 per Hubble time extrapolated to all spirals.

This indicates that a majority of nearby disk galaxies have had upto 5 Gyr (i.e. roughly corresponding to z≈0.5) to reach a quasi-stationary state. The central potential well of typical spiral galaxies seems to be deep enough to avoid strong influence from the environment. It is therefore reasonable to believe that chaotic behavior observed in the central part of most disk galaxies is due to the dynamics properties of the galaxies rather then a signature of a violent past.

## 3 Regions of Chaos

The stability of stellar orbits in disk galaxies was studied by Contopoulos [5] who used a 2D isochrone potential with a superimposed bar or spiral perturbation rotating with a constant pattern speed. For weak and intermediate bar perturbations, the main family $x_1$ of periodic orbits is stable for values of the Hamiltonian $h < h(L_1)$ where $L_1$ is the unstable Lagrangian point. Close to this point, the $x_1$ family breaks up into an infinity of families. When the perturbation is increased, the main family becomes unstable for lower values of $h$ at a bifurcation of an important resonant family and remains unstable. Typical invariant curves for orbits in a bar potential show stable regions around the families $x_1$ and $x_4$ (corresponding to retrograde orbits) representing non-periodic orbits trapped around stable periodic orbits. Chaotic regions surround the stable ones which decrease in size at stronger perturbations. The transition to chaotic motions in spiral galaxies occurs at lower values of $h$ than for bars with the same amplitude. Contopoulos [5] concluded that the two main mechanisms for producing this sudden increase of chaotic behavior are the breaking up of the $x_1$ family into an infinity of families close to $h(L_1)$, and that it becomes unstable at a resonance and remains unstable for higher $h$. The former mechanism is also responsible for the assumption that bars end just inside their co-rotation (CR) due to an increase of chaotic orbits [4].

Models for the response density in a realistic galactic potential with an imposed two armed spiral perturbation were made by Contopoulos & Grosbøl [6, 7] in the case of normal spirals and by Kaufmann & Contopoulos [15] for barred galaxies. The response was in phase and supported the imposed spiral between Inner Lindblad Resonance (ILR) and Outer Lindblad Resonance (OLR) for weak perturbations in agreement with the linear density wave theory [20]. For strong spiral potentials, the chaotic region around CR increased and the spiral was only nearly self-consistent between ILR and the 4:1 resonance (or -4:1 and OLR) due to the 45° phase shift of the stable periodic orbits just outside the 4:1 resonance. Nonlinear effects started to become important for spirals with a relative radial force perturbation $F_r \geq 5\%$ [13]. It was found that satisfactory models which matched imposed and response densities could be constructed for barred galaxies even in the regions close to CR where a significant fraction of the orbits was chaotic.

In many N-body simulation fast bars are formed while a slower rotating, more transient spiral pattern is seen outside [25]. This suggests that a spiral galaxy may have several patterns with different angular speeds. One possible explanation was given by Tagger et al. [28] who considered nonlinear coupling of such spiral modes. Bars ending at their CR could in this way be coupled to or drive outer spirals through their ILR. This mechanism provide a sharp selection of possible pattern speeds and can excite both harmonics and sub-harmonics such as m=1 waves [22]. Sellwood & Sparke [26] showed that even if bar and spiral pattern in a galaxy had different pattern speeds its appearance would most of the time suggest that the bar was connected to the inner part of the spiral as seen most frequently. Less than 10% of the time, bar and spiral in such

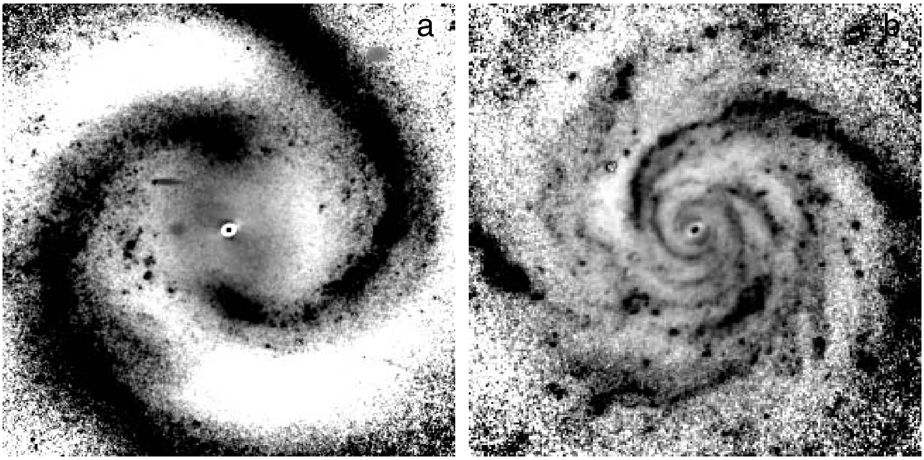



**Fig. 1.** Relative maps in the K' band of the central parts of two spiral galaxies: (**a**) NGC 1566 and (**b**) NGC 4030. Spherical bulges were subtracted before the images were de-projected. The maps were divided by their average radial profiles and presented in negative i.e. dark indicates higher than average intensity. The full range from white to black corresponds to ±30%.

systems would look fully separated such as in NGC 1566 (see Fig. 1a). Barred galaxies with rings at the end of their bar may also be candidates for systems with double pattern speeds. Although no detailed studies of the stability of orbits in the interface region between a fast bar and a slow spiral have been made, it is likely that chaotic behavior will develop there.

Whereas nonlinear response to a growing spiral perturbation will lead to increased velocity dispersion [3] and therefore possibly to more stochastic orbits, the most important locations where one would expect chaotic behaviors are the main resonances and in particular the CR region. For strong perturbation, chaos may be present for lower $h$ at an important resonance. A region of special interest is the termination of bars as it is likely to be close to CR, and in the case of an outer spiral pattern rotating with a different angular speed, be even more proven to chaotic behavior.

## 4 Chaos in the Galaxy

We have a unique opportunity to study the distribution of space velocities of individual stars in our Galaxy after the Hipparchus satellite observed proper motions and parallaxes for a large number of stars in the solar neighborhood. The local velocity distribution was recently analyzed by Dehnen [11] and Fux [12]. Velocities in the Galactic plane shows a significant structure beside its general ellipsoidal shape. Whereas features associated to stars with (B–V)<0.4 (i.e. relative young) are still likely to trace their initial conditions at formation (e.g. moving groups or open clusters), older stars will be more relaxed and can therefore be used to probe the Galactic potential. The distribution of older stars with (B–V)>0.6 displays clear velocity features like the Hercules stream which is absent in the diagram for the younger stars.

Using a realistic model of the Galaxy including a central bar, Fux [12] calculated orbits of test particles and performed 3D N-body simulation to find the possible origin of the major stream in the velocity distribution. He estimated Liapunov divergence timescales for orbits representing the local velocity distribution to quantify their stability with various parameters for the bar. Also comparing the velocity ellipsoid derived from N-body simulation, he concluded that the Hercules stream could have several origins such as being induced by cooler chaotic orbits from in the bar region or hot chaotic orbits with $h > h(\mathrm{L}_1)$. It was found possible but less likely that quasi-$\mathrm{x}_1$ orbits could contribute to streams like the Hercules and Hyades.

The analysis of the velocities in the solar neighborhood provides a detail view of a possible typical distribution in the disk of spiral galaxies. The number of stars associated with the Hercules stream could suggest that up to 15% of the stars just outside a weak bar in a disk galaxy could be chaotic. The size and shape of features like the Hercules stream also show the difficulty in observing similar structures in external galaxies due to projection effects.

## 5 Tracers of Chaos in External Galaxies

For external galaxies, one can in general not observe individual stars but only integrated properties such as surface brightness distribution and line-of-sight-velocity-profiles (LOSVP). This makes it significantly more difficult to distinguish between collections of stars following non-periodic orbits trapped around stable ones and such which exhibit a chaotic behavior. One must rely on detailed dynamic models of the galaxies to determine to what degree chaotic behavior is present in a specific region.

Chaotic regions will typically have a phase space with less structure than those occupied by stable motions although it is easy to construct models with smooth appearance consisting of stable orbits (e.g. axisymmetric disks). In disk galaxies, radial regions occupied by bars or spirals must have a high fraction of ordered motions to support such structures. If more than one non-axisymmetric mode exists in a galaxies (e.g. a bar and a spiral, or two spiral patterns), the interface between them may indicate the location of a resonance and therefore possibly a larger amount of chaotic motions than in the regions dominated by a single mode. More chaotic motions would be expected in areas between modes with different pattern speeds (e.g. in the region between the end of a bar and the start of a spiral pattern). Thus, a zone with relative small azimuthal perturbations between regions with significant bar or spiral modes would be a likely candidate for increased chaotic behavior.

A more subtle indicator for an increased fraction of chaotic motions is a radial variation of the amplitude of spiral modes. Although such variation also could be caused by the interaction of different spiral modes, their details shape may be

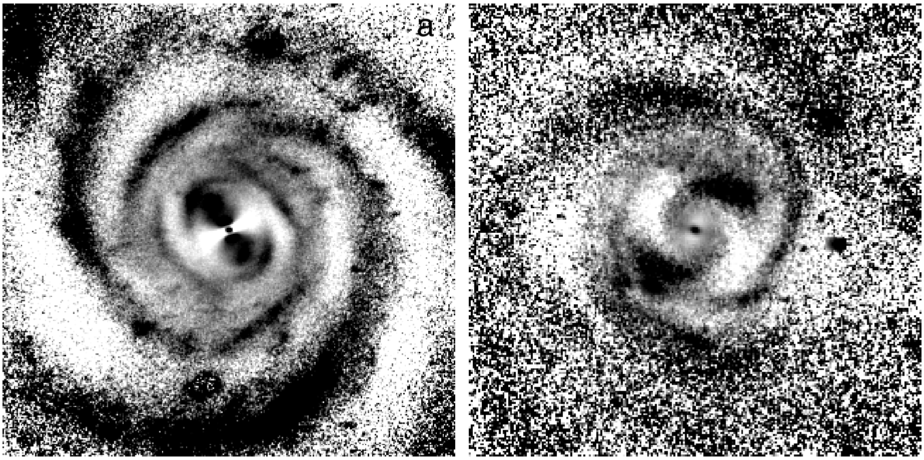



**Fig. 2.** Relative maps in the K' band of the central parts of two spiral galaxies: (**a**) NGC 4939 and (**b**) NGC 6902. The representation is identical to that of Fig. 1.

used as a diagnostic. An interaction between spiral modes would yield sinusoidal variation to the first approximation while an amplitude change due to a higher fraction of chaotic orbits in resonance regions would appear as a decrease at specific radii corresponding to major resonances in the galaxy. Attenuation of dust and patchy star formation would make it very difficult to detect such variations within the main spiral structure but radial amplitude changes in relative strong bars may display such features.

To illustration the way one may interpret the surface brightness distribution in disk galaxies, the K′ images of four spiral galaxies are shown in Figs. 1 and 2. The K band was chosen since it better represents the mass distribution of old disk stars although some population effects are still present [23]. A Sérsic $r^{1/n}$ profile [27] was fitted to the bulge and subtracted together with foreground stars before the images were de-projected. The bulge fitting was in some cases not fully satisfactory and left residuals in the very central parts. The figures show face-on, relative intensity maps normalized to the average radial profiles of the galaxies.

The first example, NGC 1566, shown in Fig. 1a, is a grand-design spiral galaxy classified as Sc(s)I in [24]. After the bulge was subtracted, a weak bar with an amplitude of ∼7% became visible. Its position angle is offset with more than 30° with respect to the start of the two armed spiral pattern and resembles the N-body simulations with different pattern speeds for bar and spiral [26]. The patchy nature of high intensity regions in the arms suggests that a significant fraction of the light in the arms originates from young objects. Especially at the start of the spiral arms just outside the bar, the star formation rate seems to be enhanced. This could be caused by gas clouds following more stochastic orbits in this region and therefore more likely collide when they encounter the spiral arms. Thus, the region between the bar and the start of the spiral is likely to exhibit some chaos.

The galaxies NGC 4030 (see Fig. 1b) has the type Sbc(r)I in [24] and show a more irregular arm structure in its inner regions while a two-armed pattern prevails at larger radii. This could indicate the existence of separate spiral modes as the inner pattern do not smoothly join the outer one. However, it is difficult to judge whether all the arm sections, seen in the central region, are associated with mass perturbations or some are mainly tracing recent star formation. If several spiral modes do exist in this galaxy, one may expect an increase of chaotic behavior where they interact.

In the case of NGC 4939 classified as a Sbc(rs)I, a strong bar is present with three sets of symmetric arcs just outside as seen on Fig. 2a. The first set of arcs is located just outside the bar but slightly offset relative to the orientation of the bar. The next set is situated almost parallel to the bar and shifted $\sim 90°$ with respect to the first arcs. Finally, a third set is again offset by $\sim 90°$ with the main grand design two-armed spiral pattern starting at the same radius. The arcs have a relative smooth appearance which suggests that they are density enhancements in the disk although significant star formation are likely to be present. The symmetry, alignment and shape of the arcs point to a stellar dynamical origin associated to specific resonances and families of periodic orbits. The exact relation can only be made after a detailed dynamic model is compared to the intensity distribution. The radial regions between the arcs (more noticeable for the two outer ones) have significantly smaller azimuthal variations than for the arcs themselves and are possibly related to a higher amount of chaotic orbits.

The last sample galaxy NGC 6902 of type Sa(r) is shown in Fig. 2b. This galaxy has two spiral pattern where the inner and outer spirals are winding with different orientations. This suggests that a major resonance is located at the radius where the two patterns join each other. An increased star formation is also observed at this location. Although the presence of spiral perturbations excludes strong chaotic behavior, an increased fraction of chaotic orbits is expected.

A more indirect way to see the results of non-linear dynamic effects and possibly increased chaotic behavior is to consider the distribution of the mean relative amplitude of the main spiral arms as function of their pitch angle for normal spirals [14] as shown in Fig. 3. It shows a lack of strong, tight spiral which could be explained by non-linear effects starting to damp growing spiral modes [3] when the relative radial force perturbation becomes large enough [13].

Two main features in the velocity distribution may be expected for regions with a substantial fraction of chaotic motion, namely: a) a general increase in the velocity dispersion and b) a non-Gaussian distribution. The chaotic orbits are not trapped around stable, periodic ones and will typically have a wider distribution function depending on the actual potential. Since it is only possible to observe one velocity component for external galaxies, the viewing angle is important as seen in the case of the velocity distribution in the solar neighborhood [11, 12] where features like the Hercules stream only could be observed at

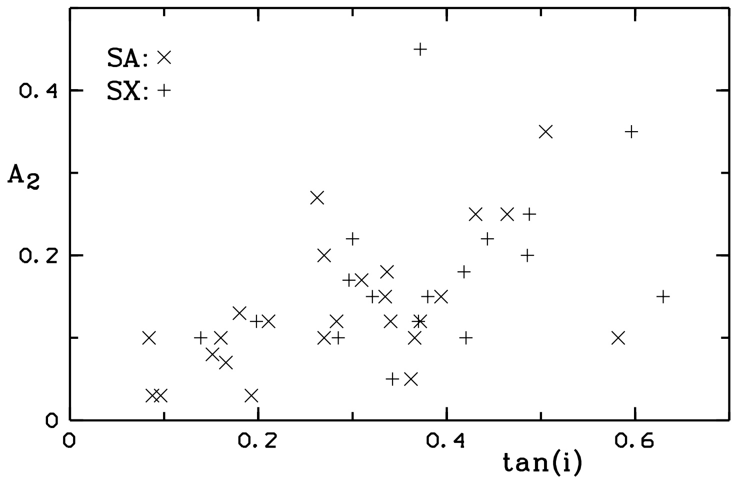



**Fig. 3.** Average relative amplitude $A_2$ of the main two-armed spiral pattern measured in K for 53 normal spiral galaxies as function of the mean pitch angle $i$ of their arms.

certain projections. Also the integration along the line of sight may mask velocity structures which originates from chaotic behavior. One may also be able to detect velocity features associated with the existence of multiple families of periodic orbits near resonance regions where chaotic motion do not dominate.

## 6 Observational Considerations

Although the study of surface brightness distribution of disk galaxies may yield some indications on possible locations of chaotic regions, it is essential to obtain detailed kinematic data in order to support a claim of chaotic behavior. It is clear from the analysis of the stellar velocity distribution in the solar neighborhood [12] that a unique interpretation may be very difficult even with high quality data.

The need for an accurate dynamic model demands that both the mass distribution in the disk and the total potential including a possible dark matter component are estimated. The main problems in deriving a mass distribution from surface photometry are population effects and attenuation by dust. These effects are significantly reduced when using the near-infrared K band [23] as can be seen in Fig. 4 where both B and K maps of NGC 2997 are shown. Strong dust lanes are seen in the B band along the major arms but also in the inter-arm regions. Further, and the bulge appears significantly more prominent in the K band. Although attenuation by dust is strongly reduced in K, population effects must be considered as one notices strings of knots along the arms. Their compactness and location close to the main dust lanes suggest that they are associated to young objects (e.g. star forming regions). Besides these knots, it is

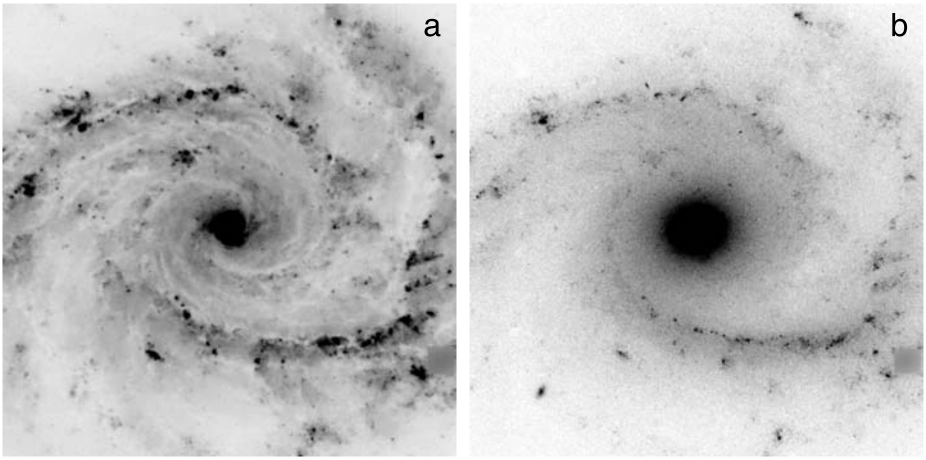



**Fig. 4.** Images of NGC 2997 observed in **a**) B band image and (**b**) K′ band. Foreground stars were removed.

likely that a more defuse component of young stars contributes to the K band luminosity in the arm regions.

The rotation curve of a galaxy can be obtained through long slit spectroscopy (LSS) along the major axis or using an integral field unit (IFU) which yields a full velocity map. It is simpler to used emission lines to measure the velocity field, however, since they measure the gas kinematics corrections for possible effects due to shocks, streaming motions and differences in velocity dispersion compared to the stellar component must be applied. A safer approach is to measure stellar absorption lines (e.g. MgI at 518 nm or CaII at 854 nm). They still have to be corrected for velocity perturbations in the disk (e.g. spiral or bar modes) before the average potential can be derived. Systematic effects due to attenuation by dust [1] and asymmetric, non-Gaussian velocity profiles [18] should also be considered.

It is also important to choose the region suspected to exhibit chaotic behavior carefully including its position relative to its parent galaxy. If it is close to the major axis of the galaxy, it is necessary to subtract the velocity component due to the general rotation of the galaxy. At the minor axis, a contribution to the LOSVP from the bulge may be significant. In all cases, the integration over the finite thickness of the disk will introduce systematic effects.

Taking as an example the local stellar velocity distribution, one would have to detect velocity features with a separation of ∼50 km/sec and an amplitude contrast of less than 10%. This would require a signal-to-noise ratio (SNR) in the range of 20-50 depending of number of free parameters in the model.

## 7 Feasibility of Observing Chaos

Whereas it is trivial to obtain both deep K band surface photometry and long slit spectra for deriving a general potential model of a disk galaxy with 4m class telescopes, the observation of detailed LOSVP's with sufficient SNR and spectral resolution is significantly more challenging. To estimate the feasibility of such observations with current state-of-the-art instrumentation, the ESO Very Large Telescope (VLT) facility was taken as an example. Its four 8m unit telescopes are located at Paranal in the Atacama desert, Chile, and provide excellent conditions for this type of project. At present, four VLT instruments could be considered for obtaining LOSVP, namely:

**FORS1/2** have both imaging and spectroscopic modes in the visual part of the spectrum. The maximum spectral resolution is $\sim$1700 for long slit mode.

**VIMOS** is a visual multiple object spectrograph with imaging modes. It has several IFU modes including one with a field of almost 1 arcmin$^2$ and a spectral resolution of $\sim$2200.

**FLAMES/GIRAFFE** is a multi-fiber, high resolution spectrograph for visual wavelengths. There are several small IFU's and one $7' \times 11'$ IFU head with a lower resolution of $\sim$9000.

**ISAAC** is an infrared instrument with both imaging and long slit spectroscopic modes. It's higher spectral resolution in the K band is $\sim$3000 which is just sufficient to resolve the OH lines and thereby give access to the low background inter-line regions.

The typical surface brightness at the end of a bar in a spiral galaxy is K $\approx$ 18 mag/arcsec$^2$ with a color index (I-K) $\approx$ 2 mag. Strong spiral arms are on average at least 1 mag. fainter while inter-arm regions typically are 1 mag. fainter than the arms. To estimate the feasibility of observing chaos with the VLT, the ESO Exposure Time Calculator(version 2.8.3) [2] was used assuming a surface brightness in K = 18 mag/arcsec$^2$ or I = 20 mag/arcsec$^2$ with a seeing of $0.8''$ and an airmass of 1.2. The results are given in Table 1 where spectral resolution and typical SNR are listed for a 1 hour exposure. The SNR estimate for ISAAC may vary significantly depending on the exact location of the lines to be measured relative to OH lines.

By averaging over several spectral and spatial channels, a somewhat higher SNR can be obtained. Even so, it is clear from the estimates in Table 1 that it is just feasible to obtain an acceptable SNR and spectral resolution at the end of the bar while regions in the spiral structure would require even larger facilities.

## 8 Conclusions

Studies of the past star formation rate and the morphology of galaxies in the HDF's indicate that a majority of local disk galaxies was formed at least 5 Gyr ago and therefore has had time enough to reach a relaxed, quasi-stable state. There are no evidence that bar and spiral structure depend on the environment

**Table 1.** Performance of VLT instruments in spectroscopy mode for a 1 hour exposure with 0.8″ seeing. The wavelength ($\Delta\lambda$) and velocity ($\Delta$v) resolution per detector pixel together with the Point Spread Function (PSF) are given for each configuration. Finally, the number of electron from the source and the corresponding SNR are listed as calculated by the ESO Exposure Time Calculator for an extended source with a surface brightness of I = 20 mag/arcsec$^2$.

| Instrument | Mode | | $\Delta\lambda$ (nm/pix) | $\Delta$v (km/s/pix) | PSF (pix) | Source ($e^-$) | SNR |
|---|---|---|---|---|---|---|---|
| FORS | LSS | 600I | 0.132 | 50 | 4 | 329 | 27 |
| | LSS | 1200R | 0.075 | 34 | 4 | 274 | 26 |
| | LSS | 1400V | 0.063 | 36 | 4 | 154 | 15 |
| ISAAC | LSS | MR | 0.121 | 16 | 7 | 195 | 3: |
| VIMOS | IFU | R2150 | 0.061 | 29 | 5 | 248 | 10 |
| | IFU | R1000 | 0.273 | 116 | 5 | 1531 | 31 |
| GIRAFFE | IFU | LR04 | 0.020 | 11 | 3 | 62 | 5 |

of the parent galaxy. Thus, chaotic behavior observed in disk galaxies is likely to have an internal dynamic origin if there is no evidence of recent mergers.

The most likely regions to find chaos in spiral galaxies are the major stellar resonances in the disk especially CR. If a galaxy hosts several spiral modes with different angular speeds (e.g. a fast bar and a slower rotating spiral), one would expect increased chaotic behavior in the interface region between them.

The analysis of the stellar velocity distribution in the solar neighborhood show that some streams (e.g. Hercules) may originate from a population of stars with chaotic orbits. Even with access to high quality data for individual stars, the interpretation is ambiguous and rely on detailed comparisons with dynamic models.

For external galaxies where only integrated properties can be observed, it is possible to identify regions where chaotic behavior may be expected but it is essential to compare detailed LOSVP with dynamic model to access the probability of chaos. Candidate regions are major resonances and interface zones between different spiral modes. It is important to consider possible contamination of measured velocity profiles by galactic rotation, disk thickness and attenuation by dust.

Whereas surface photometry and basic kinematics data can be obtained with 4m class telescopes, velocity profiles with sufficient spectral resolution and SNR are much more demanding. With current state-of-the-art facilities like VLT, it is just feasible to access a surface brightness of I $\approx$ 20 mag/arcsec$^2$ corresponding to the end of the bar in a typical disk galaxy. Observation to search for chaotic behavior in the disk related to the main spiral structure would only be possible with significantly larger facilities.

## References


1. M. Baes, H. Dejonghe: Mon. Not. R. Astron. Soc. **335**, 441 (2002)
2. P. Ballester, A. Disarò, A. Dorigo et al.: ESO Messenger **96**, 19 (1999)
3. Z. Bin, Y. Zeng-yuan: Appl. Math. Mech. (Eng. Ed.) **11**, 901 (1990)
4. G. Contopoulos: Astrophys. Astron. **81**, 198 (1980)
5. G. Contopoulos: Astrophys. Astron. **117**, 89 (1983)
6. G. Contopoulos, P. Grosbøl: Astrophys. Astron. **155**, 11 (1986)
7. G. Contopoulos, P. Grosbøl: Astrophys. Astron. **197**, 83 (1988)
8. S. van den Bergh: Astron. J. **124**, 782 (2002)
9. S. van den Bergh: Astron. J. **124**, 786 (2002)
10. S. van den Bergh: Pub. Astron. Soc. Pacific **114**, 797 (2002)
11. W. Dehnen: Astron. J. **119**, 800 (2000)
12. R. Fux: Astrophys. Astron. **373**, 511 (2001)
13. P. Grosbøl: Pub. Astron. Soc. Pacific **105**, 651 (1993)
14. P. Grosbøl, E. Pompei, P.A. Patsis: ASP Conf. Ser. **275**, eds. E.Athanassoula, A.Bosma, R.Mujica, 305 (Astron. Soc. Pacific, San Francisco 2002)
15. D.E. Kaufmann, G. Contopoulos: Astrophys. Astron. **309**, 381 (1996)
16. W. Keel, W. Wu: Astron. J. **110**, 129 (1995)
17. S.M. Kent: Astron. J. **91**, 1301 (1986)
18. K. Kuijken, M.R. Merrifield: Mon. Not. R. Astron. Soc. **264**, 712 (1993)
19. K.M. Lanzetta, N. Yahata, S. Pascarelle et al.: Astrophys. J. **570**, 492 (2002)
20. C.C. Lin, F.H. Shu: Astrophys. J. **140**, 646 (1964)
21. P. Madau, L. Pozzetti, M. Dickinson: Astrophys. J. **498**, 106 (1998)
22. F. Masset, M. Tagger: Astrophys. Astron. **322**, 442 (1997)
23. H.-W. Rix, M.J. Rieke: Astrophys. J. **418**, 123 (1993)
24. A. Sandage, G.A. Tammann: *A Revised Shapley-Ames Catalog of Bright Galaxies* (Carnegie Inst., Washington 1981)
25. J.A. Sellwood: Mon. Not. R. Astron. Soc. **217**, 127 (1985)
26. J.A. Sellwood, L.S. Sparke: Mon. Not. R. Astron. Soc. **231**, 25P (1988)
27. J.L. Sérsic: *Atlas de galaxias australes* (Obs. Astron. de Cordoba, Cordoba 1968)
28. M. Tagger, J.F. Sygnet, E. Athanassoula, R. Pellat: Astrophys. J. **318**, L43 (1987)
29. N. Voglis, G. Contopoulos, C. Efthymiopoulos: Cel. Mech. Dyn. Astron. **73**, 211 (1999)

# Observational Determination of the Gravitational Potential and Pattern Speed in Strongly Barred Galaxies

Per A.B. Lindblad and Per Olof Lindblad

Stockholm Observatory, AlbaNova, SE–10691 Stockholm, Sweden

**Abstract.** In order to compute stellar orbits in spiral galaxies the gravitational potential and its pattern speed must be known. Observationally, these parameters are difficult to determine, in particular for strongly barred galaxies. We will briefly review different methods and illustrate in more detail the case where the problem has been approached by numerical gasdynamical simulations.

## 1 Introduction

When computing orbits in real galaxies one needs to know the gravitational potential and its pattern speed, i.e. the angular velocity of its non-axisymmetric component. In the quasi-steady density wave picture of galaxy dynamics these two quantities are considered to be more or less constant throughout the system and over a certain period of time.

To derive the gravitational potential from photometry requires very accurate multicolour photometry to very faint levels with corrections for extinction and assuming mass-luminosity ratios for a mixture of stellar populations. To this must be added the potential of an unknown amount of dark matter. Where the symmetry plane of the galaxy is suitably inclined to the plane of the sky, radial velocities will give information of the kinematics in the plane of the galaxy, complementing the information from photometry.

The axisymmetric part of the potential in the plane of the galaxy is generally described by the rotation curve, whereby we mean the set of circular orbital velocities as a function of the distance from the centre, given by the axisymmetric Fourier component of the density distribution.

The interplay between the orbital motions and the pattern speed gives rise to various resonance phenomena, of which we in particular note the corotation resonance (CR) and the Inner (ILR) and Outer (OLR) Lindblad Resonances. The CR occurs where the orbital angular velocity is close to the pattern speed. Any regular non-circular particle orbit can be described as a closed orbit with $n$ pericentra, rotating with a certain angular velocity. The rotation can be slower than the mean orbital velocity, so that the particle describes the rotating closed orbit mainly in the forward direction, or faster, so that the particle describes the closed orbit mainly in the retrograde direction. Where the rotation of the orbit, for $n = 2$, is close to the pattern speed, we have an ILR in the former case and an OLR in the latter. For one and the same pattern ILR and OLR fall

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 213–222, 2003.

on different sides of CR. There remains the possibility that one and the same system may contain different structural components each with its own pattern speed [10].

In the case of a weak bar with small deviations from circular motion we have a circular corotation region, where the circular angular velocity is close to the pattern speed. This case is well demonstrated by England, Hunter and Contopoulos [2] by hydrodynamical model computations of motions in various bar potentials. Figure 1 shows the case of a rather weak bar perturbation. To the left is shown the gas density response pattern when the gas has been settled into a quasi-steady state, and to the right the gas velocity vectors in a frame rotating with the pattern.

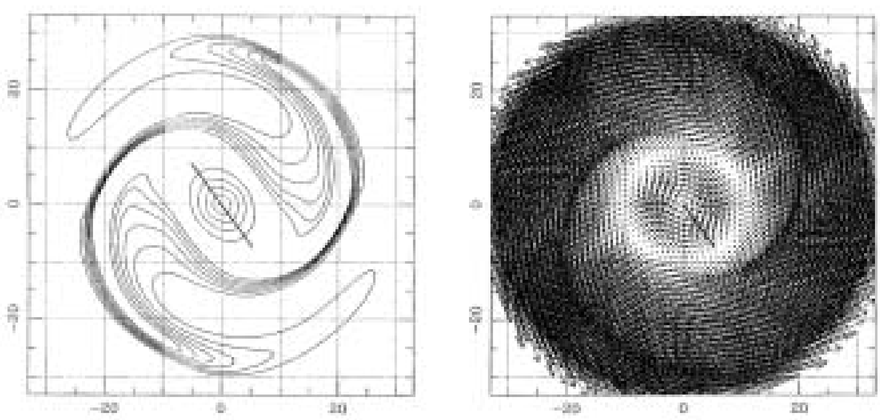


**Fig. 1.** Gas density response (left) and velocity vectors (right) of a rotating bar perturbation of moderate strength. The system rotates clock-wise. A straight line in the figures shows the position and extent of the bar. From [2]

The circular corotation region of low velocities is clearly seen in this figure. It contains two vortex regions, corresponding to the Lagrangian L4 and L5 equilibrium points of the restricted 3-body problem. In these regions the gas rotates counter clock-wise and expands from the vortex centre.

The case of a very strong bar, as shown in Fig. 2, however, is different. In the bar region there is no continuous set of nearly circular orbits and no circle of corotation resonance. Here the corotation region has broken down. There are still the two vortices placed at right angels to the bar, now very pronounced. In addition, there are two vortices close to the gas density maxima in the bar, around which the gas now flows in the same direction as the bar rotation and inwards towards the vortex centre. There also seem to be low velocity corotation regions outside the ends of the bar. The entire bar is close to corotating.

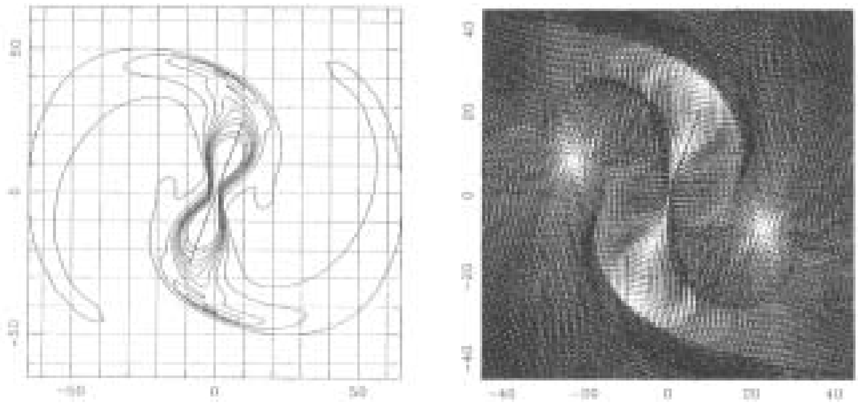


**Fig. 2.** Gas density response (left) and velocity vectors (right) of a rotating very strong bar perturbation. The system rotates clock-wise. A straight line in the figures shows the position and extent of the bar. From [2]

To derive a rotation curve in the bar region from the observed radial velocities in such a galaxy would be a very difficult task and its usefulness might be questioned.

## 2 Methods to Determine the Pattern Speed

A sophisticated scheme to derive the pattern speed for a non-circular-symmetric pattern in open spiral and barred galaxies was suggested by Tremaine and Weinberg [11]. The scheme is based on the continuity equation and is rather model independent. It is assumed that the disk has a well defined pattern speed, that the surface brightness of the tracer obeys the continuity equation, and that there is no streaming velocity normal to the disk plane.

The continuity equation is integrated over a strip parallel to the apparent major axis of the system. The gain and loss of matter across the strip, due to the rotation of the pattern, is related to the radial velocities and luminosity distribution along the strip.

Figure 3 illustrates the version of this method designed by Merrifield and Kuijken [9], here applied to the early type SB galaxy NGC 4596 [3]. The spectra are added all along the slit, and the right side shows the Doppler broadening function of this single absorption line spectrum determined by means of a templet star. This is done for three different slit offsets along the minor axis. On the bottom we see the continuum luminosity distribution along the same slits. Each pair gives a point in the diagram. The slope of the line is proportional to the pattern speed multiplied with the sin $i$ of the inclination of the plane of the galaxy to the plane of the sky. This can be extended to any number of slits.

The method has been applied to a handful of early type barred galaxies. Its potential usefulness, however, for H I observations in barred galaxies is doubtful. Figure 4 shows an optical image as well as the total neutral hydrogen map of the

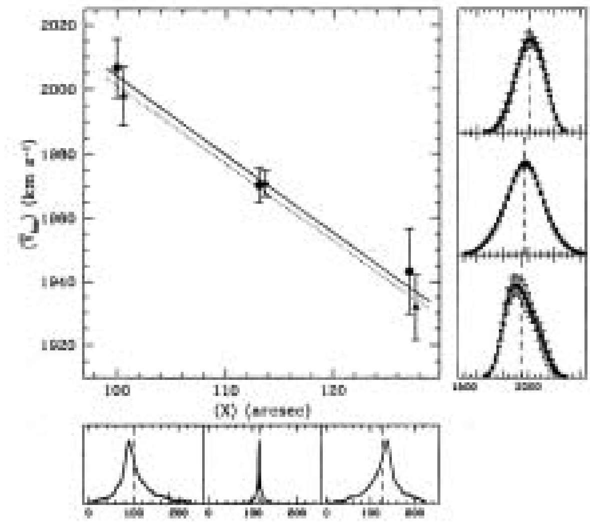



**Fig. 3.** Mean line-of-sight velocities versus luminosity centroid position for three slits parallel to the major axis of NGC 4596. From [3]

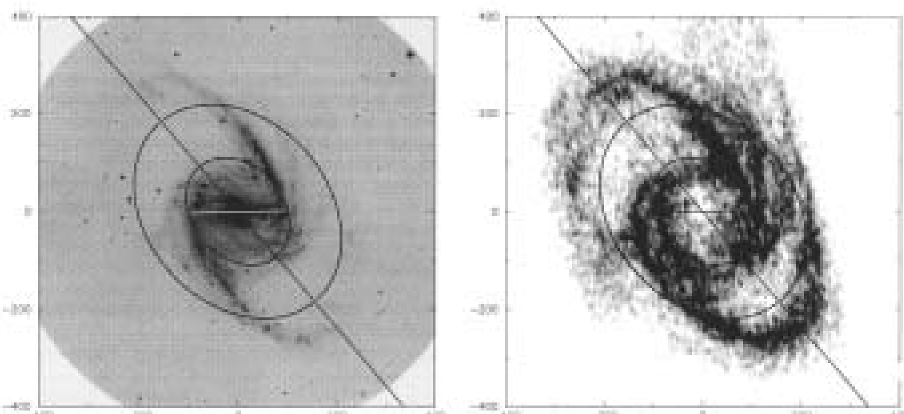


**Fig. 4.** An optical ESO 3.6 m prime focus plate (left) and total H I column density map (right) of NGC 1365. The ellipses enclose the region where the rotation curve is believed to give a reliable description of the axisymmetric forces. The bar major axis and the line of nodes are marked as straight lines running through the centre. From [5]

barred galaxy NGC 1365. Typically, H I disappears in the bar region, where the deviation from circular symmetry is the largest and the method is most useful. The continuity equation is not valid. In the spiral arms H I is forming stars and along the bar H I is transferred to molecular hydrogen streaming towards the centre. Our efforts have shown that the method cannot be meaningfully applied to the H I data for NGC 1365.

Several authors have tried to identify the positions of a corotation circle, or other specific resonances, in real galaxies and from the rotation curve get the pattern speed. In some cases this procedure seems to start from assumptions that should be proven in the end.

A method, based on the linear density wave theory, has been suggested by Canzian [1]. According to this theory the residual radial velocities along the line of sight, when the rotation curve has been subtracted, should be proportional to $\sin \Phi$ close to the ILR, where $\Phi$ is the central angle in the plane of the galaxy, and proportional to $\sin 3\Phi$ close to the OLR. At corotation the amplitudes of the two functions should have a specific ratio, dependent of the shape of the rotation curve. The method is very sensitive to the proper derivation of the rotation curve and the residual velocities. Being based on the linear density wave theory it should not be applicable to barred spirals.

If there is no easy way out, the ultimate method is to make a simulation of the entire galaxy, taking into consideration all available observational information. Several authors have done this in different ways. In the course of such a simulation procedure one should get a rotation curve, perturbing potential and pattern speed, consistent with the observed velocity field and giving resonance regions compatible with the observed morphology.

## 3 The Case of NGC 1365

To illustrate the procedure, let us consider the case of NGC 1365 which was simulated by Per A.B. Lindblad in a project at Stockholm Observatory, in which we collaborated with E. Athanassoula in Marseille [5]. A similar analysis has been performed, among others, by Weiner, Sellwood, Williams, and van Gorkom in the case of the SB galaxy NGC 4123 [13] [12], using the same gasdynamical code.

NGC 1365 (Fig. 4) is one of the more thoroughly studied nearby isolated barred galaxies [6]. Its inclination to the plane of the sky of 40° is suitable for radial velocity studies of the kinematics, and the inclination of the bar to the line of nodes close to optimal for a study of streaming both across and along the bar. The distance is 18 Mpc, which gives a scale where 1″ corresponds to 100 pc. With a diameter of 11′, or 66 kpc, it is a supergiant galaxy.

Detailed VLA observations in H I have been presented by Jörsäter and van Moorsel [4]. The total H I density map is given in Fig. 4. As was mentioned, H I is very scarce in the bar region. In the very nucleus H I is seen in absorption.

However, the central region was filled in with velocities from long slit emission line spectra, and a complete radial velocity map for the interstellar gas was constructed [7]. In Fig. 5 we see the characteristic twist along the bar and wiggles along spiral arms.

In Figs. 4 and 5 the ellipses separate the *bar region*, *intermediate region* and *outer region*. As the basis for the rotation curve we adopt the azimuthally averaged rotation curve of Jörsäter and van Moorsel. However, in the bar region

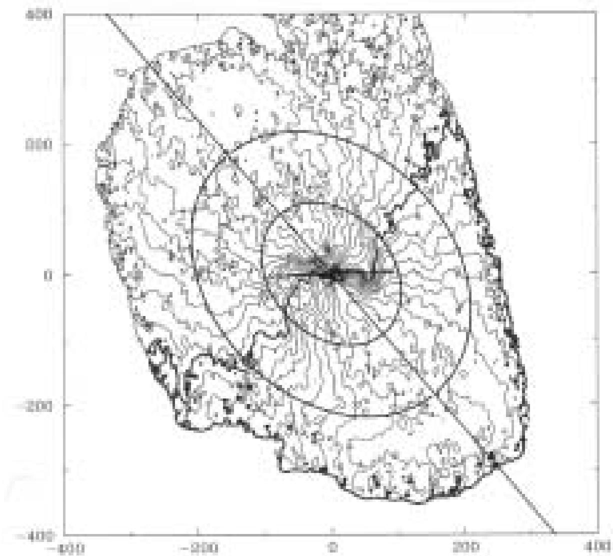



**Fig. 5.** The observed radial velocity field of NGC 1365. The contour interval is 20 km/s, and the systemic velocity is drawn as a thick line. The ellipses are the same as in Fig. 4. From [5]

such an azimuthally averaging method does not give a good approximation to the rotation curve, and in the outer region the system is warped, as seen from the kinematics in Fig. 5, which again makes the rotation curve uncertain.

To estimate the perturbing potential, we use infrared photometry, as the bar region shows a multitude of dust. We choose an infrared J-band image obtained as part of the Ohio-State University Bright Galaxy Imaging Survey (Fig. 6). This image was analysed in terms of even azimuthal Fourier components. As the spiral arms seem firmly attached to the ends of the bar, we assume that the spiral part of the structure, at least for a considerable time, has the same pattern speed as the bar.

At the start of the fitting procedure we let the axisymmetric part of the potential be represented by the Jörsäter–van Moorsel rotation curve. The perturbing potential is derived from the infrared surface photometry, where the mass/luminosity ratio $M/L$ (in arbitrary units, as the photometry is not absolutely calibrated) is kept as a free parameter. This free parameter also compensates for effects of the unknown thickness of the bar.

The model fit in the intermediate region is rather insensitive to the exact shape of the rotation curve in the bar region. Thus, we can now make simulations with a sequence of different pattern speeds and $M/L$ ratios to get the best fit to the structure in the *intermediate region*. The detailed structure in this region is particularly sensitive to the choice of pattern speed, and this speed can now be fixed to within a few km s$^{-1}$ kpc$^{-1}$, at a value of 18 km s$^{-1}$ kpc$^{-1}$ for NGC 1365.

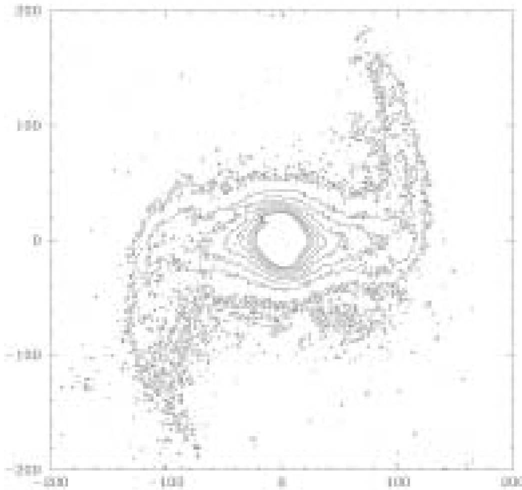


**Fig. 6.** Inclination corrected J-band image of NGC 1365. From [5]

With the chosen value for the pattern speed we now adjust the Jörsäter–van Moorsel rotation curve in the *bar region* as well as the final choice of the perturbing $M/L$ ratio, until the simulation reproduces the observed radial velocities from slit spectra as well as the morphology of the dust lanes along the bar, both of which lay constraints on the position of the ILR.

The result is seen in Fig. 7, where density contours of the model are overlaid the H I total column density map. The slight mismatch in the outer region is due to the steep decline of the Jörsäter–van Moorsel rotation curve. If raised about 10 km/s in the outer region, the OLR and spiral arms move outward (i.e. a less drastic warp is assumed) and the match is improved.

Figure 8 compares the observed radial velocity field with that given by the model as observed from the same angle. The forced bisymmetry of the model limits the fit in both Figs. 7 and 8. In spite of this, the main features are reproduced fairly well. The velocity pattern in Fig. 8 shows the characteristics of orbits elongated along the bar. The shocks along the bar are smoothed in frame (**a**) due to the procedure with which the velocity field was constructed from randomly positioned slits. Thus, here a comparison should be made directly with slit measurements. Figure 9 shows this for a slit placed perpendicular to the bar 29″ East of the nucleus. The jump across the shock of 300 km/s is well reproduced.

Thus, we arrive at a possible variation of the axisymmetric forces and a pattern speed which are in agreement with the observed morphology and velocity field. In contrast, Weiner et al. [12] through accurate photometry and by the perturbations required, assuming the same $M/L$ for disk and bar, derive by simulations similar to ours this $M/L$ and deduce the mass of the halo required to reproduce the outer rotation curve. This means that they use the perturbations from the bar to infer a $M/L$ for the disk and bar and get the mass of the dark

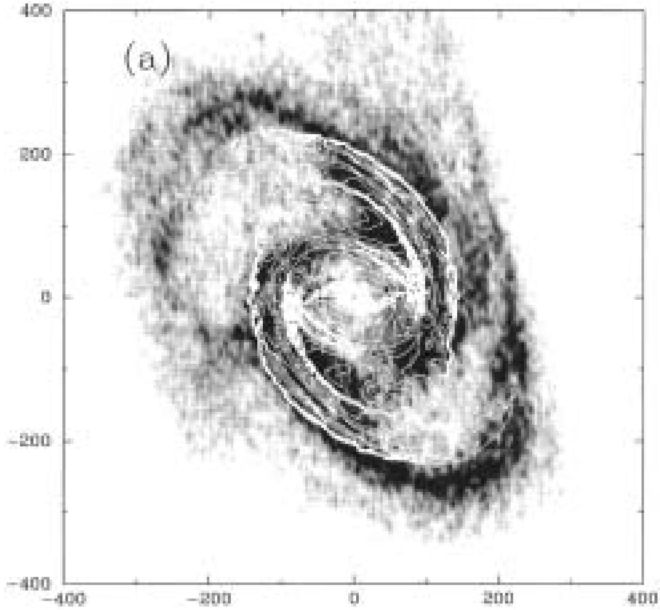



**Fig. 7.** Model density contours overlaid the total column H I density map. From [5]

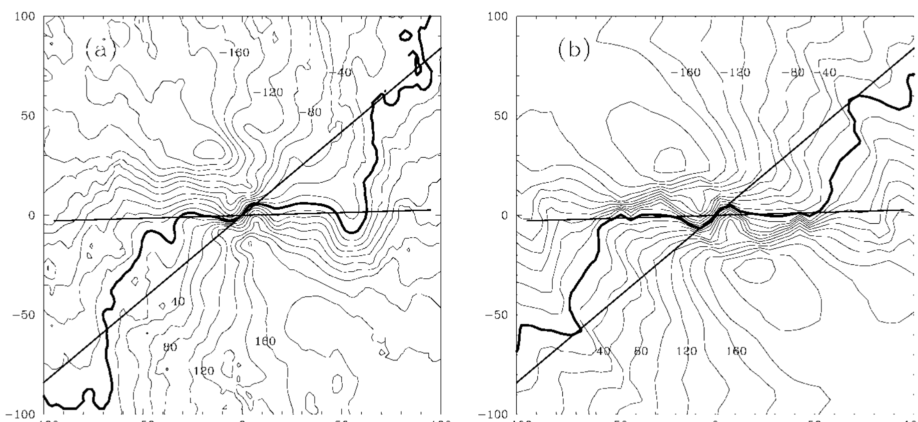



**Fig. 8.** Comparison between the observed velocity field in NGC 1365 (**a**) and that given by the model (**b**). The orientations of the bar axis in PA 92° and the minor axis of the galaxy in PA 130° are shown as straight lines. From [5]

halo component. For NGC 1365 this would not be safe due to the warp, which makes the outer rotation curve uncertain.

The gas flow in a frame rotating with the pattern is shown in Fig. 10, where the circles mark the ILR, CR and OLR resonance positions. Only half of the velocity field is shown. The orbits show the familiar twist around the ILR [8] as well as the change of direction of the flow at the shock fronts along the bar. The

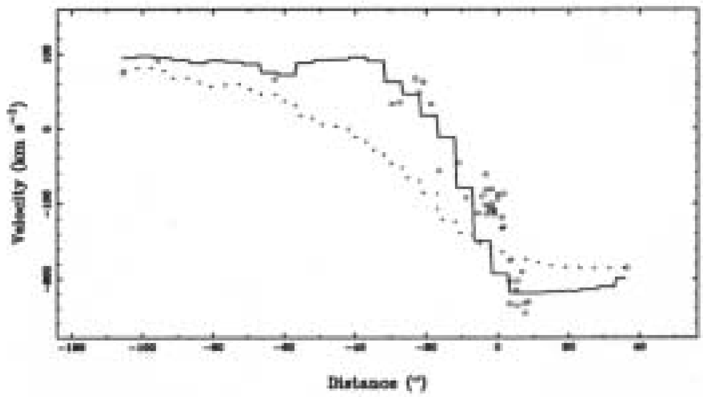



**Fig. 9.** Radial velocities along a slit placed perpendicular to the bar 29″ East of the nucleus. Open circles: observed velocities. Solid line: model velocities. Dotted line: pure rotational motion according to the rotation curve. From [5]

spiral arms extending from the ends of the bar appear also for a purely barred perturbation, but the spiral part of the potential is necessary to drive the arms through corotation. The vortex regions at CR are apparent.

Thus, through detailed comparison between observations and models, estimates of the pattern speed as well as total gravitational potential can be obtained for individual galaxies. This will permit us to compute stellar orbits in rather realistic, but in our case bi-symmetrical, galaxy potentials.

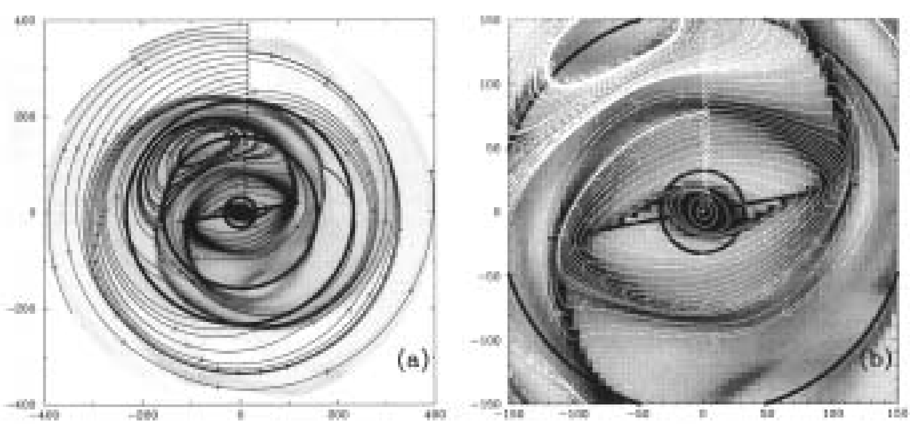



**Fig. 10.** Gas flow lines from the model overlaid gray scale maps of the same model. Full scale (**a**), bar region (**b**). From [5]

## References


1. B. Canzian: Astrophys. J. **414**, 487 (1993)
2. M.N. England, J.H. Hunter Jr., G. Contopoulos: Astrophys. J. **540**, 154 (2000)
3. J. Gerssen, K. Kuiken, M.R. Merrifield: Month. Not. Roy. Astr. Soc. **306**, 926 (1999)
4. S. Jörsäter, G. van Moorsel: Astron. J. **110**, 2037 (1995)
5. P.A.B. Lindblad, P.O. Lindblad, E. Athanassoula: Astron. Astrophys. **313**, 65 (1996)
6. P.O. Lindblad: Astron. Astrophys. Rev. **9**, 221 (1999)
7. P.O. Lindblad, M. Hjelm, J. Högbom, S. Jörsäter, P.A.B. Lindblad, M. Santos-Lleó: Astron. Astrophys. Suppl. **120**, 403 (1996)
8. P.O. Lindblad, P.A.B. Lindblad: Publ. Astr. Soc. Pacific Conf. **66**, 29 (1994)
9. M.R. Merrifield, K. Kuijken: Month. Not. Roy. Astr. Soc. **274**, 933 (1995)
10. J.A. Sellwood, L.S. Sparke: Month. Not. Roy. Astr. Soc. **231**, 25p (1988)
11. S. Tremaine, M.D. Weinberg: Astrophys. J. **282**, L5 (1984)
12. B.J. Weiner, J.A. Sellwood, T.B. Williams: Astrophys. J. **546**, 931 (2001)
13. B.J. Weiner, T.B. Williams, J.H. van Gorkom, J.A. Sellwood: Astrophys. J. **546**, 916 (2001)

# Observational Manifestation of Chaos in Grand Design Spiral Galaxies

Alexei M. Fridman[1,2], Roald Z. Sagdeev[3],
Oleg V. Khoruzii[1,4], and Evgenii V. Polyachenko[1]

[1] Institute of Astronomy RAS, 48 Pyatnitskaya st., Moscow 117647, Russia
[2] Sternberg Astronomical Institute, Moscow State University, University prospect, 13, Moscow, 119899, Russia
[3] University of Maryland, College Park, MD. 20742, USA
[4] Troitsk Institute for Innovation and Thermonuclear Research, National Research Center of Russian Federation, Troitsk, Moscow Reg. 142092, Russia

**Abstract.** To study dynamic properties of the gaseous disk of the grand design spiral galaxy NGC 3631 we calculate the Lyapunov characteristic numbers (LCN) for different families of streamlines in the disk. For the trajectories near separatrices of the giant vortices and near saddle points presenting in the velocity field, the LCN turned out to be positive. The result is insensitive to the method of the calculation. Both methods — using two trajectories and based on linearized equations — give the identical results. The values of the LCN in the gaseous disk of NGC 3631 are independent on the Riemannian metric used for the calculations in agreement with the classical mathematical theorem. The spectra of the 'short-time' LCN (stretching numbers) evaluated for the same trajectories turned out to be non-invariant. We confirmed this result obtained for the real galactic disk on classical model examples.

## 1 Introduction

The main topic of our paper consists in the demonstration of a fact that results of analysis of the observed velocity field of the galactic disk can serve as a source of our knowledge of the stochastic galactic dynamics.

This topic has relation to a gaseous disk rather than to stellar one. In spite of the evident recent progress in the measurements of the line-of-sight velocity field of stellar disks, our knowledge of stochastic stellar dynamics of external galaxies is based for the most part on theoretical investigations. Information in this respect from observational data on stellar velocity fields is moderate for the following reason.

As a rule, the external galaxies are not resolved into individual stars. Using the analysis of absorption lines we can measure the line-of-sight velocity field of the stellar population averaged over some local spatial region. The trajectory of this region of the stellar disk may differ qualitatively from the trajectories of the stars in the same region. For example, a stellar bar rotates as a solid body, while the stellar orbits in the bar may be complicated and far from a simple rotation. Such a behaviour is typical for collisionless selfgravitating systems.

From observational data we can construct gravitational potential as a function of coordinates. But some variations of the potential within errors of obser-

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 223–234, 2003.

vations often result in the transformation of regular stellar orbits into chaotic ones and vice versa. This may lead to artefacts.

At first glance the line-of-sight velocity field of a gaseous disk also can not be used directly to study the stochastic dynamics of the disk. However, the use of the method of restoration of 3D velocity field from the observed line-of-sight velocity field [1], [2] enables to determine regular and chaotic trajectories by the calculation of the Lyapunov characteristic numbers (LCN) [3].

In this case a natural question may appear. It is well known (see, e.g. [4]), that the LCN is calculated for the trajectories in the phase space, while the restoration method [1], [2] gives 3D velocity field in the coordinate space[1]. Hence it allows to see the behaviour of the trajectories in the coordinate space rather than in the phase one. The book [5] may help to resolve this question. In this book the hydrodynamical equations of 3D stationary incompressible flows are reduced to nonstationary dynamical equations in 2D phase space. In other words, in this book it is shown, that the problem of analysis of the properties of trajectories in 3D stationary incompressible flows is equivalent to the problem for nonstationary dynamical systems with 2D phase space. These systems, evidently, can demonstrate both regular and chaotic motions. The same is done in [6] for compressible 3D stationary flows. The idea of the reduction is the following.

A steady-state 3D flow is described by the set of equations:

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v_x(x,\, y,\, z)\,, \quad \frac{\mathrm{d}y}{\mathrm{d}t} = v_y(x,\, y,\, z)\,, \quad \frac{\mathrm{d}z}{\mathrm{d}t} = v_z(x,\, y,\, z)\,, \tag{1}$$

which can be rewritten in the following form:

$$\frac{\mathrm{d}x}{v_x} = \frac{\mathrm{d}y}{v_y} = \frac{\mathrm{d}z}{v_z} = \mathrm{d}t\,. \tag{2}$$

For our aim, a more convenient notation is

$$\frac{\mathrm{d}x}{\mathrm{d}z} = \frac{v_x}{v_z} \equiv f_1(x,\, y,\, z)\,, \quad \frac{\mathrm{d}y}{\mathrm{d}z} = \frac{v_y}{v_z} \equiv f_2(x,\, y,\, z)\,. \tag{3}$$

The latter equations show that we are dealing with the "nonstationary" problem for a dynamical system in 2D phase space $(x,\, y)$. The variable $z$ is playing the role of time $\tau$:

$$\frac{\mathrm{d}x}{\mathrm{d}\tau} = f_1(x,\, y,\, \tau)\,, \quad \frac{\mathrm{d}y}{\mathrm{d}\tau} = f_2(x,\, y,\, \tau)\,. \tag{4}$$

These equations describe the dynamical systems where stable and unstable trajectories may coexist. As it follows from above (and [5], [6]), that corresponds to the coexistence of stable and unstable streamlines in 3D coordinate space.

[1] In more details see [1]–[3].

## 2 Restored Velocity Field of the Grand Design Spiral Galaxy NGC 3631

In Fig. 1 one can see the reconstructed velocity field of NGC 3631[2] in the galactic plane with superimposed lines of constant phase of the vertical (perpendicular to the galactic plane) velocity. Squares mark the maxima of the absolute values of the vertical velocity of gas at each radius. Asterisks show the locations of the zeros of the vertical velocities. The vertical velocity amplitude is not shown. Thin lines mark the location of the vortices — anticyclones (upper left and lower right) and cyclones.

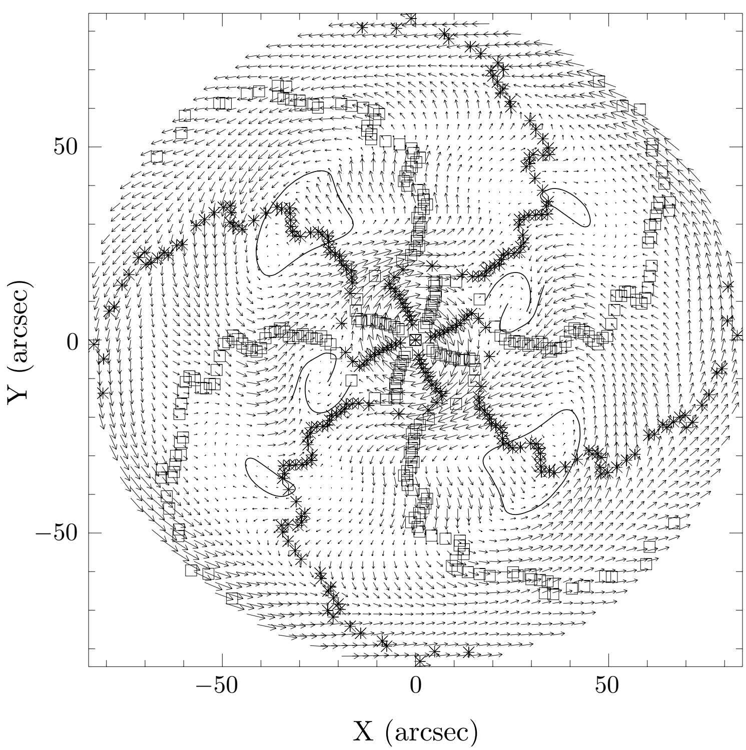



**Fig. 1.** The reconstructed velocity field of NGC 3631 in the galactic plane and superimposed lines of constant phases of the vertical motions. Squares mark the maxima of the absolute values of the vertical velocity of gas at each radius. Asterisks show the locations of the zeros of the vertical velocities. Thin lines show the location of the vortices — anticyclones (upper left and lower right) and cyclones.

[2] A solution of the ill-posed problem of the reconstruction of three component velocity field from the observed line-of-sight velocity field of gaseous disk of grand design spiral galaxies is described in papers [1], [2], [7], [8] and in the review [9].

From observations it follows that gaseous disks of spiral galaxies – the set of their main parameters – lie close to the boundary of their dynamic instability [10]. That might seem natural: in the course of developing the instability, the velocity dispersion grows, and the disk approaches the boundary of the instability [9]. Since the instability generating both the spiral arms and vortices is saturated, then the 3D motion of the gas should be quasi-stationary in the reference frame corotating with the density wave. It means that trajectories coincide with streamlines in 3D coordinate space. In the regions close to those where $V_z$ equals to zero, every fluid particle participates in 2D motion only and hence its trajectory should coincide with its 2D streamlines.

If vortex lie in the regions of the 2D motion, the fluid particles trapped in the vortex are separated from the transiting (untrapped) ones by the separatrix – the last closed streamline around the vortex center. As we can see in Fig. 1, the vortices with centers close to the zeros of the vertical motions ($V_z = 0$) are surrounded by closed separatrices. Two cyclones located far from the zeros do not demonstrate the presence of a clear separatrix in the 2D streamlines. These facts can be considered as an evidence of the real three-dimensionality of the velocity field of the gaseous disk. The whole structure of the reconstructed velocity field in Fig. 1 agrees with the assumption of its quasi-stationarity.

Besides the separatrices surrounding vortices the 2D velocity field in the disk plane contains saddle points (marked by crosses in Fig. 2). Choosing the beginning of streamlines near separatrices or near the saddle points, one can see that these streamlines diverge. We would like to know, if this divergence is exponential or not, in other words, if the streamlines are chaotic or regular. To clarify it we need to calculate the LCN.

## 3 The Calculation of the Lyapunov Characteristic Numbers

In the case of exponential divergence of the trajectories we have

$$d(t) \sim d_0 e^{\lambda t} , \tag{5}$$

where $d_0$ is the initial separation between neighbouring trajectories, $d(t)$ is the separation for the time $t$, $\lambda$ is a rate of the exponential divergence and is equal to the maximum LCN.

The rigorous definition of the LCN is [4]

$$\lambda = \lim_{t \to \infty,\, d_0 \to 0} \frac{1}{t} \ln \frac{d(t)}{d_0} . \tag{6}$$

In our case of the gaseous disk of NGC 3631, it is difficult to use the definition of the LCN (6) for the following reasons:

1) the duration of observations is much smaller than the characteristic time $\lambda^{-1}$ of the exponential divergence of two points moving along nearby trajectories;

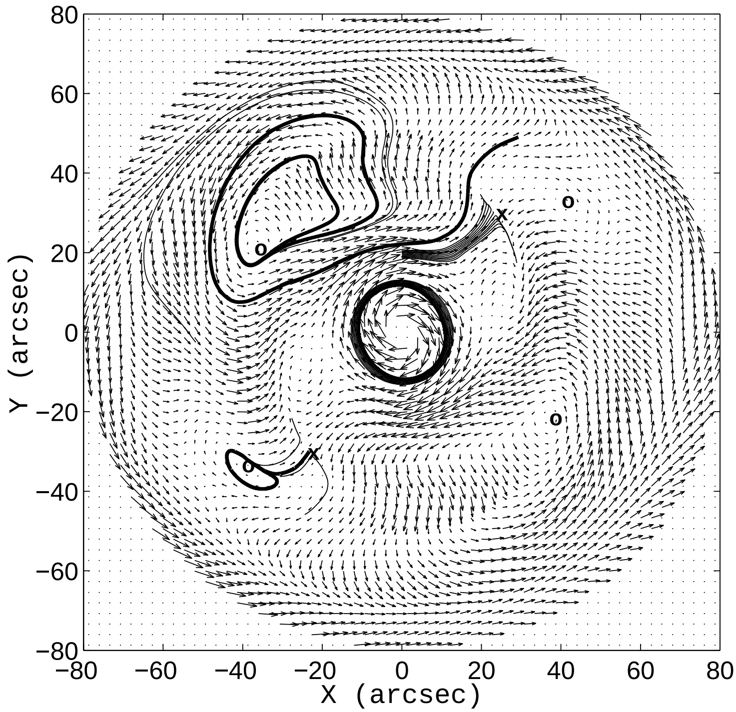



**Fig. 2.** The divergent trajectories near the vortices separatrices and a set of trajectories near the saddle point superimposed on the restored two dimensional velocity field. Thick solid lines show the trajectories which are used to calculate the stochastic characteristics in the vicinity of the vortices. "o"-signs mark the centers of the vortices. "x"-signs mark the saddle points. Also, non-divergent trajectories near the center of the disk are shown.

2) the presence of a minimal distance $d_{\min}$ between two trajectories owing to a finite spatial resolution of measurements $\delta$, $d_{\min} \geq \delta$;
3) the ratio of the characteristic scale $R_{ch}$ of the velocity field variations to the resolution $\delta$ is not too large, moreover there are some regions where $R_{ch} \simeq \delta$;
4) we cannot measure the velocity field of the overall disk but only of a part of the disk.

To overcome the first difficulty we use the mentioned above property of the stationarity of the velocity field.

The second and the third difficulties restrict an allowable maximal length of the trajectory (and thus a maximal time $T$ of the calculation of the LCN). First, a trajectory can eventually leave the area, for which the velocity field is defined. Second, a trajectory may come to the region where the characteristic scale of the velocity field variations is of the order of the spatial resolution of the velocity data, $R_{ch} \simeq \delta$ (according to the second restriction), that contradict to the linear

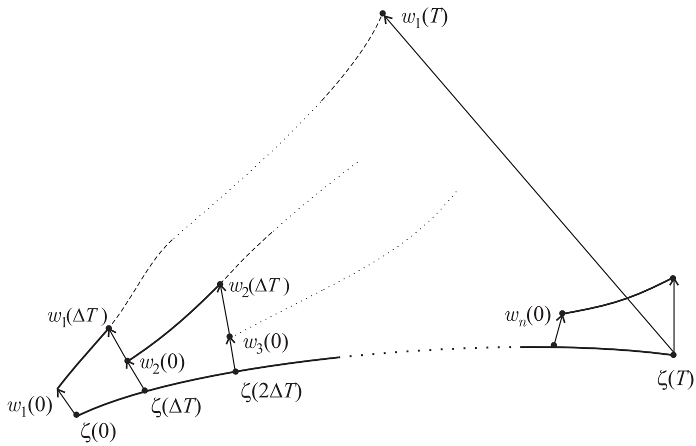



**Fig. 3.** The method of the LCN calculation. The reference trajectory is denoted by the longest solid line and some auxiliary trajectories by dashed curves with short regions of solid curves.

approximation condition $d/R_{ch} \ll 1$. Hence, these regions are "forbidden" for the method used to find the LCN.

In the previous paper [3] (hereafter Paper I) we tried to overcome the difficulty, connected with the limitation on the maximal length of the trajectory, using a method proposed by Casartelli *et al.* [11] and described in the well-known monograph [4] (see also [12], [13], [14]).

In our case we specify the number of steps $n$ and divide the given time interval $[0, T]$ into $n$ time intervals $\Delta T = T/n$. Choosing the initial deviation vector $\boldsymbol{w}_1(0)$ we evaluate two trajectories from $\boldsymbol{\zeta}_0$ and $\boldsymbol{\zeta}_0 + \boldsymbol{w}_1(0)$ and determine $\boldsymbol{w}_1(\Delta T)$ (see Fig. 3). After each step, following [11], we will renormalize the deviation vector to the initial length $d_0$, preserving its direction

$$\begin{aligned} \boldsymbol{\zeta}_{i+1}(0) &= \boldsymbol{\zeta}(i\Delta T) + \boldsymbol{w}_{i+1}(0), \\ \boldsymbol{w}_{i+1}(0) &= \frac{d_0}{d_i}\boldsymbol{w}_i(\Delta T), \end{aligned} \tag{7}$$

where $d_i \equiv d(\boldsymbol{w}_i(\Delta T))$.

According to [11] and [15], for sufficiently large $T$ the reliable estimate of the LCN (6) is the following:

$$\lambda \approx \lambda^{(n)} = \frac{1}{T}\sum_{i=1}^{n} \ln \frac{d_i}{d_0}. \tag{8}$$

The described method implies integration over two trajectories — the reference trajectory, shown in Fig. 3 by the thick solid line, and an auxiliary trajectory.

The limited length of the trajectory used in the calculations poses a question on the accuracy of the LCN determination. We consider the results to be reliable when:

$$\xi \equiv \lambda^{(n)} T \gg 1\,. \tag{9}$$

Sometimes, the trajectories are so short that the condition (9) is not fulfilled. According to [13] and [14] the reliability can be improved, if one takes a set of trajectories instead of one. In this case the LCN is calculated as follows:

$$\lambda^{(n)} = \frac{1}{N} \sum_{k=1}^{N} \lambda_k^{(n)}\,, \tag{10}$$

where $\lambda_k^{(n)}$ is calculated according to the formula (8), using the $k$-th trajectory as a reference one, $N$ is the total number of the basic trajectories. The reliability condition turns into

$$\eta \equiv \xi N = \lambda^{(n)} T N \gg 1\,. \tag{11}$$

All details of the calculations of the LCN for different families of streamlines in the gaseous disk of NGC 3631 are contained in Paper I. Our calculations of the LCN led to the conclusion, that the gaseous disk of NGC 3631 contained both the regular and chaotic streamlines. The formers are located near the disk center, the latters – in the vicinity of separatrices of vortices and near the saddle points.

To perform the calculations of the LCN one needs to define the Riemannian metric $d$. The general form of the metric used in the present work is

$$d(\boldsymbol{w}) = \sqrt{g_1 x^2 + g_2 y^2}, \tag{12}$$

where $x$ and $y$ are the Cartesian coordinates of the vector $\boldsymbol{w}$. Metrics $d$ differ by the positive metric coefficients $g_1$ and $g_2$. For the sake of simplicity we refer to different metrics as follows: $(g_1, g_2)$. For example, we often use metric (1,1), that implies $d = \sqrt{x^2 + y^2}$.

The Oseledec theorem [16] claims that the result of the LCN calculation should not depend on the Riemannian metric. It would be interesting to note that for the considered real system the LCN, calculated using different metrics, also deviate very little (see Fig. 4). Namely, near the separatrix of the anticyclone for the metric (1,1) $\lambda^{(n)} = 0.8768 T_0^{-1}$, for the metric (1,0) $\lambda^{(n)} = 0.8931 T_0^{-1}$, for the metric (0,1) $\lambda^{(n)} = 0.8513 T_0^{-1}$ ($T_0 = 7.5 \cdot 10^7$ years)[3].

In parallel with the use of the mentioned above technique, which employs two trajectories for evaluation of the LCN, we recalculated LCN for the same regions using the linearized equation for the deviation vector $w$ (see [4]). It turns out that for the sufficiently smooth vector field interpolation[4] the results obtained in both cases are identical.

[3] According to the metric definition if $x^2$ or $y^2$ is positive, we should have $d > 0$. Rigorously speaking, metrics $(g_1, g_2) = (1{,}0)$ and $(0{,}1)$ do not fulfil this requirement. To overcome this obstacle one can assume that the corresponding coefficients have infinitely small positive values.

[4] Here we use fourth order tensor product spline interpolation.

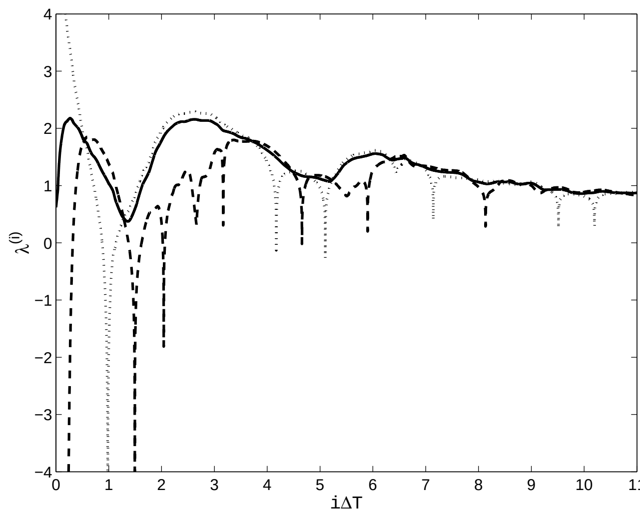



**Fig. 4.** The behavior of $\lambda^{(i)}$ with $i\Delta T/T_0$ for trajectory near the separatrix of the anticyclone, calculated for different metrics: (1,1) – solid line, (1,0) – dashed line, (0,1) – dotted line. Number of time steps $n = 1000$.

## 4 Spectrum of the Stretching Numbers

The expansion coefficient (in terminology of Oseledec [16]) or the stretching number [17] [13] [14] is defined as follows:

$$a_i = \frac{1}{\Delta T} \ln \frac{d_i(\Delta T)}{d_0} \, . \tag{13}$$

Using (13) one can find the spectra of the stretching numbers for real objects. According to the definition [13], [14], the spectrum of the stretching numbers is

$$S(a, x_0, y_0) = \lim_{N \to \infty} \frac{1}{N} \frac{\mathrm{d}N(a)}{\mathrm{d}a} \, , \tag{14}$$

where $\mathrm{d}N(a)$ is the number of appearances of the stretching number $a_i$ in the interval $(a, \, a + da)$.

The spectra, calculated for chaotic and regular trajectories using different metrics are shown in Figs. 5–8. In all cases, the spectra are not invariant to the metric change.

This is very interesting fact, since as it follows from (8), (13) and (14) the LCN is the first moment of the spectrum of the stretching numbers

$$\lambda = \int aS(a)da, \tag{15}$$

but the LCN itself preserve the mentioned invariance.

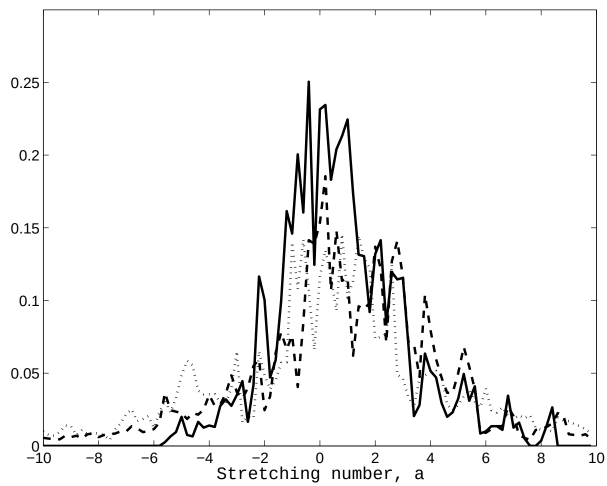



**Fig. 5.** Spectra for the stochastic trajectory in the vicinity of the anticyclone (see Fig. 2), calculated for different metrics: (1,1) – solid line, (1,0) – dashed line, (0,1) – dotted line. Number of time steps $n = 1000$.

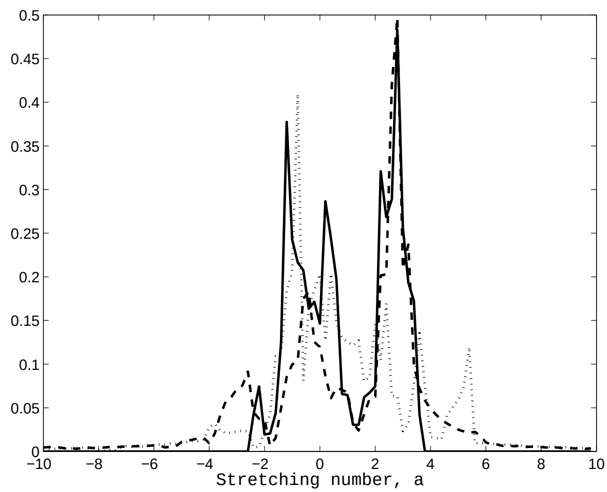



**Fig. 6.** Spectra for the stochastic trajectory in the vicinity of the cyclone (see Fig. 2), calculated for different metrics: (1,1) – solid line, (1,0) – dashed line, (0,1) – dotted line. Number of time steps $n = 1000$.

The dependence of the spectrum form on the metric has not been known previously, although it can be checked that it holds either in theoretical and real observable systems. Figures 9, 10 show the LCN and the spectra, calculated for different metrics for the well-known standard (Chirikov) map and the Lorenz attractor. One can see that in both cases the LCN, calculated using different metrics are equal, whereas the forms of the spectra are different.

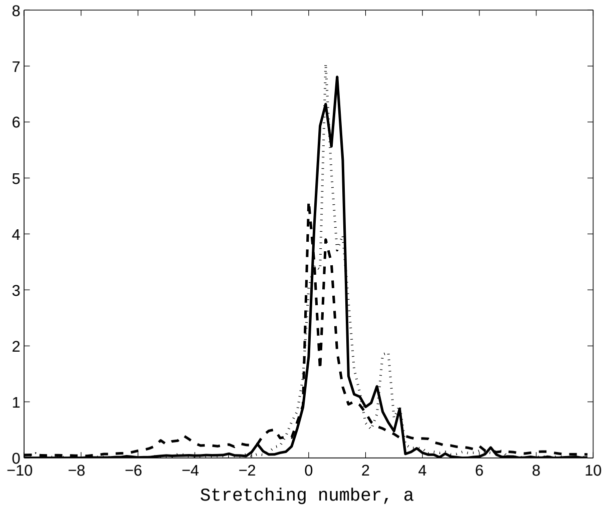



**Fig. 7.** Spectra for the set of stochastic trajectories near the saddle point (see Fig. 2), calculated for different metrics: (1,1) – solid line, (1,0) – dashed line, (0,1) – dotted line. Number of time steps $n = 1000$.

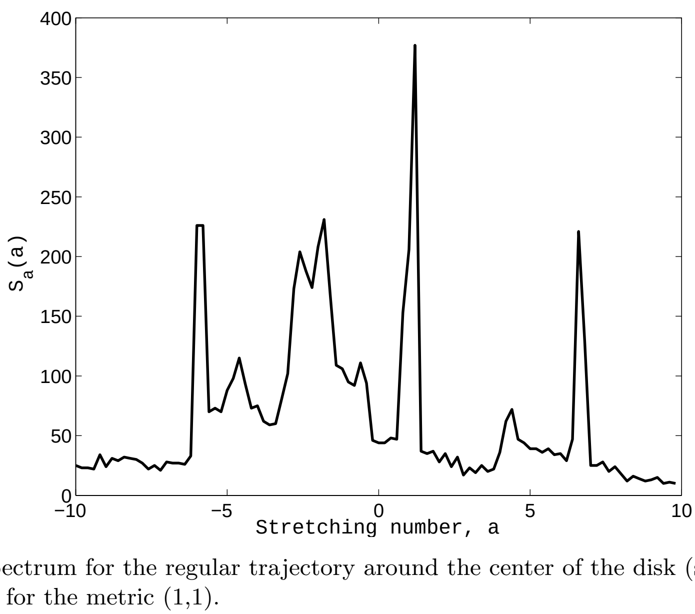



**Fig. 8.** Spectrum for the regular trajectory around the center of the disk (see Fig. 2), calculated for the metric (1,1).

## 5 Conclusions

In conclusion let us summarize the dynamical properties of the gaseous disk of NGC 3631 revealed in this paper and the Paper I.

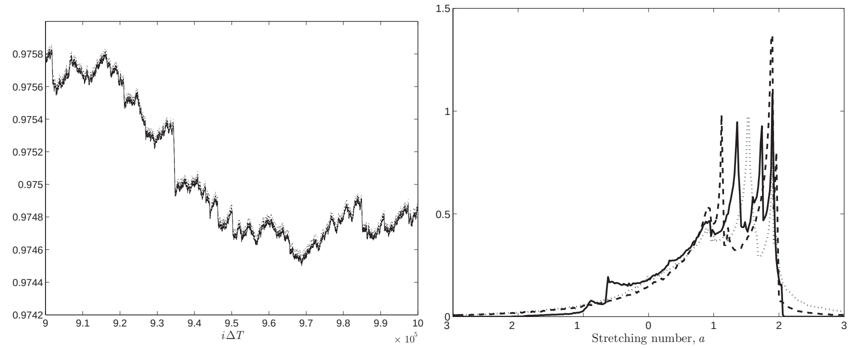



**Fig. 9.** The LCN (left figure) and the spectra (right figure) for the standard (Chirikov) model, calculated for different metrics: (1,1) – solid line, (0,1) – dashed line, (1,0) – dotted line.

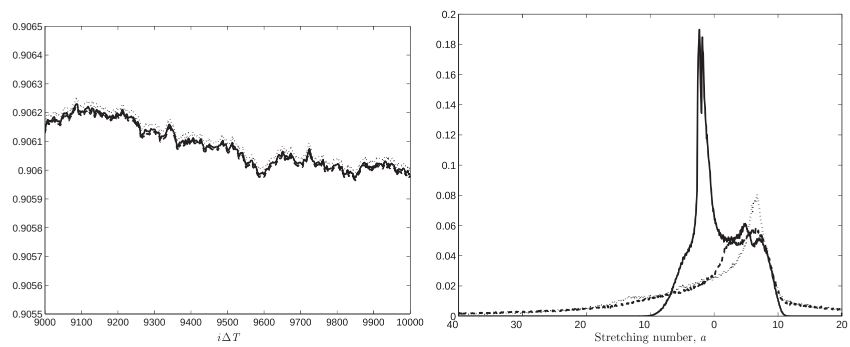



**Fig. 10.** The LCN (left figure) and the spectra (right figure) for the Lorenz attractor) model, calculated for different metrics: (1,1) – solid line, (0,1) – dashed line, (1,0) – dotted line.

1. The three component velocity field restored from the observed line-of-sight velocity field of the gaseous disk of the galaxy NGC 3631 is stationary and demonstrates the presence of both regular and stochastic trajectories of the gas.
2. The regular trajectories are observed near the center of the disk, the stochastic ones — near the saddle points and the separatrices of giant vortices presenting in the velocity field.
3. The type of the divergence of the trajectories was determined by two different methods of the calculation of the Lyapunov characteristic numbers (LCN): using two neighbouring trajectories and based on the linearised equations. Both methods gave identical results.
4. The LCN obtained for the real galaxy are turn out to be invariant to the metric change, in full agreement with the Oseledec mathematical theorem [16].
5. For the first time it was demonstrated, that the form of the spectra of the 'short-time' LCN (stretching numbers) varies with the metric change.

6. The correctness of two latter conclusions was confirmed also for the classical model dynamical systems — the standard map and the Lorenz attractor.

**Acknowledgments**

Authors are grateful to V. Arnold, B. Chirikov, G. Contopolous, Y. Fridman, V. Oseledec, M. Rabinovich, and Ja. Sinai for numerous and very fruitful discussions. This work was performed with partial financial support of RFBR grants N 02-02-16878, N 02-02-06603, "Leading Scientific Schools" grant N 00-15-96528, and "Fundamental Space Researches. Astronomy" grants: N 1.2.3.1, N 1.7.4.3.

## References

1. V.V. Lyakhovich, A.M. Fridman, O.V. Khoruzhii, A.I. Pavlov: Astronomy Report **41**, 447 (1997)
2. A.M. Fridman, O.V. Khoruzhii, V.V. Lyakhovich, V.S. Avedisova, O.K. Sil'chenko, A.V. Zasov, A.S. Rastorguev, V.L. Afanasiev, S.N. Dodonov, J. Boulesteix: Astroph. and Space Sci. **252**, 115 (1997)
3. A.M. Fridman, O.V. Khoruzhii, E.V. Polyachenko: 'Observational Manifestation of Chaos in the Gaseous Disk of the Grand-Design Spiral Galaxy NGC 3631'. In: *Observational Manifestation of Chaos in Astrophysical Objects, International Workshop, Moscow, August 28-29, 2001*, ed. by A.M. Fridman, M.Ya. Marov, R.H. Miller, Space Science Rev. **102**, 51 (2002)
4. A.J. Lichtenberg, M.A. Lieberman: *Regular and Stochastic Motion* (Springer-Verlag, New York, Heidelberg, Berlin 1983)
5. G.M. Zaslavsky, R.Z. Sagdeev, D.A. Usikov, A.A. Chernikov: *Weak Chaos and Quasi-Regular Patterns* (Cambridge Univ. Press, Cambridge 1991)
6. Govorukhin et al. (1999)
7. A.M. Fridman, O.V. Khoruzhii, V.V. Lyakhovich, O.K. Sil'chenko, A.V. Zasov, V.L. Afanasiev, S.N. Dodonov, J. Boulesteix: Astron. & Astroph. **371**, 538 (2001)
8. A.M. Fridman, O.V. Khoruzhii, E.V. Polyachenko, A.V. Zasov, O.K. Sil'chenko, A.V. Moiseev, A.N. Burlak, V.L. Afanasiev, S.N. Dodonov, J.H. Knapen: Mon. Not. R. Astr. Soc. **323**, 651 (2001)
9. A.M. Fridman, O.V. Khoruzhii: Space Science Rev., accepted for publication in 2002.
10. A.V. Zasov, S. Simakov: Astrofizika **29**, 190 (1988)
11. M. Casartelli, E. Diana, L. Galgani, A. Scotti: Phys. Rev. A **13**, 1921 (1976)
12. B.V. Chirikov, F.M. Izrailev, V.A. Tayurski: Comput. Physics Commun. **5**, 11 (1973)
13. N. Voglis, G. Contopoulos: J. Phys. **A27**, 4899 (1994)
14. G. Contopoulos, N. Voglis: A&A, 317, 73 (1997)
15. G. Bennetin, L. Galgani, J.-M. Strelcyn: Phys. Rev. **A14**, 2338 (1976)
16. V.I. Oseledec: Tr. Mosk. Mat. Obsch., **19**, 179, 1968 (Trans. Mosc. Math. Soc.**19**, 197, 1968).
17. C. Froeschle', Ch. Froeschle, E. Lohinger: Cel. Mech. Dyn. Astron. **56**, 307 (1993)

# Quarter-Turn Spirals in Galaxies

Evgenii Polyachenko

Institute of Astronomy RAS, 48 Pyatnitskaya st., Moscow 117647, Russia

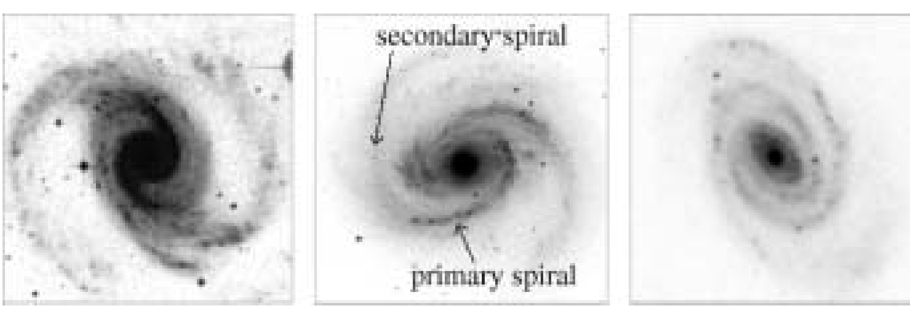
**Abstract.** Observations in the optical show that grand design spirals consist of a set of principle arms and characteristic near-circular extensions that can be described as quarter-turn spirals. Arguments are presented in favor of the idea that the latter set of spirals is caused by the response of the material of the disk to the gravitational potential of the main spiral arms. The peculiarities of the potential in the narrow transitional annulus between the regions of spiral and multipole behavior can explain basic characteristic features of the quarter-turn spirals (their angular length and small pitch angles).

## 1 Introduction

Optical observations give many examples of grand design spiral galaxies in which arms consist of two parts: strong symmetric primary spirals, and adjacent faint secondary spirals. Figure 1 shows three images of such galaxies. The transition between the parts is marked by steep brightness gradients and by changing in the pitch angle of the arms.

As it is shown in many papers, for many spiral galaxies the last part of the optical grand design spirals almost vanishes in the near infrared wavelengths (see, e.g., [1], [2], [3]). Such a discrepancy between optical and near IR data suggests possible different formation mechanisms of these parts and allow one to consider the secondary spirals as a specific part of a whole spiral structure.

Basic characteristic features of the secondary spirals can be established by studying the azimuthal Fourier spectra of brightness maps for such galaxies:




NGC 1566 NGC 4321 NGC 5364

**Fig. 1.** Examples of galaxies with QTS

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 235–242, 2003.

1. Their pitch angles are small (compared to those of the primary spirals);
2. Their angular length is of about 90°.

Due to the first two features one may refer to the secondary arms as the quarter-turn spirals (QTS). Lack of the QTS in the near IR implies that

3. The secondary spirals formed mainly from the cold component of the galactic disk.

Below, the explanation of the QTS phenomena is presented.

## 2 The Simplest Theory

In this section the simplest theory is described, in which QTS are treated as the disk response to the gravitational potential of the primary spirals. The calculations are carried out under the following basic assumptions:

- the galaxies under consideration have well defined spiral structures with certain pattern velocities;
- the simplest approximation of a cold 2-dimensional disk can be used (it is expected that this approximation contains the main effects, and the corrections for the velocity dispersion are small);
- linear theory of the disk response can be used;
- QTS are located sufficiently far from the main resonances.

Accordingly, the density response to the gravitational potential can be calculated by employing the linearized hydrodynamical equations with the pressure equal to zero

$$\frac{\partial \tilde{\sigma}}{\partial t} + \Omega \frac{\partial \tilde{\sigma}}{\partial \varphi} + \frac{1}{r}\frac{\partial}{\partial r}(r\sigma_0 v_{r1}) + \frac{\sigma_0}{r}\frac{\partial v_{\varphi 1}}{\partial \varphi} = 0,$$

$$\frac{\partial v_{r1}}{\partial t} + \Omega \frac{\partial v_{r1}}{\partial \varphi} - 2\Omega v_{\varphi 1} = -\frac{\partial \Phi_1}{\partial r},$$

$$\frac{\partial v_{\varphi 1}}{\partial t} + \Omega \frac{\partial v_{\varphi 1}}{\partial \varphi} + \frac{\kappa^2}{2\Omega} v_{r1} = -\frac{1}{r}\frac{\partial \Phi_1}{\partial \varphi},$$

where $v_{r1}$ and $v_{\varphi 1}$ — the perturbed velocities, $\kappa$ is the epicyclic frequency, $\kappa^2 = 4\Omega^2 + r d\Omega^2/dr$, $\Omega = \Omega(r)$ is the disk angular velocity, $\tilde{\sigma}$ and $\Phi_1$ are the disk (density) response and the potential of the primary spirals, respectively. Assuming that all perturbations are proportional to the exponent $e^{i(m\varphi - \omega t)}$, one can obtain the response in the form

$$\tilde{\sigma} = -\frac{1}{r}\frac{\partial}{\partial r}\left(r\varepsilon \frac{\partial \Phi_1}{\partial r}\right) + \varepsilon \frac{m^2}{r^2}\Phi_1 + \frac{2m}{r\omega_*}\frac{\partial}{\partial r}(\varepsilon \Omega)\Phi_1, \tag{1}$$

where

$$\varepsilon \equiv \frac{\sigma_0(r)}{\omega_*^2 - \kappa^2} \tag{2}$$

is the gravitational analog of dielectric permittivity [4], [5]; $\omega_* = \omega - m\Omega(r)$; $\omega = m\Omega_p$ is the perturbation frequency, $\Omega_p$ is the pattern speed.

### 2.1 Qualitative Considerations

In the case when $\Phi_1$ varies quickly with radius (it is proved below by numerical calculations), one can neglect all but one term in (1) with the highest derivative of $\Phi_1$. Thus, formula (1) turns to

$$\tilde{\sigma} \simeq -\varepsilon \Phi_1''.$$

It is also implied here that the contributions from the resonance terms is of no significance. As it is clear from (2), $\varepsilon(r) < 0$ in the region between the inner and outer Lindblad resonances, at which $\Omega_p = \Omega - \kappa/m$ and $\Omega_p = \Omega + \kappa/m$, respectively. As a rule, the spiral structure is localized within this region (see, e.g. [6]). Then we obtain the qualitative formula for the response in the form:

$$\tilde{\sigma} \propto \Phi_1''. \tag{3}$$

To use the formula (3) one need to know the potential of the primary spirals. In the general case, this potential should be calculated numerically. However, the behavior of the potential can be predicted from qualitative considerations in two regions: in the spiral region and in the region sufficiently far from the spirals.

**1. Spiral Region.** For tightly-wound spirals, the curves of minima of the potential coincide with the curves of maxima of the surface density of the primary spirals. It follows from the well-known relation between the potential $\Phi_1$ and the surface density $\sigma_1$ in Toomre's theory of the tightly-wound spirals [7]:

$$\Phi_1(r) = -\frac{2\pi G}{|k(r)|}\sigma_1(r),$$

where $k(r)$ is the wavenumber, $G$ is the gravitational constant. The same correspondence approximately holds for open spirals, as it is demonstrated below on the model examples (see also [8]).

Applying the formula (3) for such a spiral-like potential $\Phi_1(r) \propto e^{-iF_\Phi}$ (here the phase $F_\Phi = \int^r k(r')dr'$), one can obtain

$$\tilde{\sigma}(r) \propto \Phi_1''(r) \simeq -k^2 \Phi_1(r), \tag{4}$$

i.e. in this region, the curves of maxima of the density response follow the curves of minima of the gravitational potential (see Fig. 2).

**2. Region Beyond the Primary Spirals.** Well away from the spirals, the potential $\Phi_1$ tends to its asymptotic quadrupole form [9]

$$\Phi_1(r,\varphi) \to -r^{-3}\cos 2(\varphi - \varphi_0), \quad r \to \infty, \quad \varphi_0 = const. \tag{5}$$

Applying the formula (3) to the quadrupole potential (5) one can obtain:

$$\tilde{\sigma}(r) \propto \Phi_1''(r) \simeq +12\Phi_1(r), \tag{6}$$

i.e. the curves of maxima of the response follow the curves of maxima of the gravitational potential (see Fig. 2).

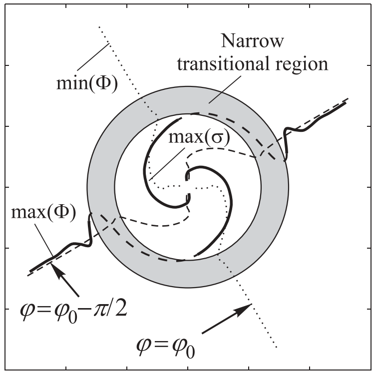



**Fig. 2.** Spiral, multipole, and transitional behavior of the gravitational potential.

It is notable, that the curves of the potential minima take its asymptotic direction $\varphi = \varphi_0$ just after departing from the primary spirals near its ends. It follows from numerical calculations of the potential. This implies that the effective region, that determines the asymptotic direction of the curves of the potential minima is small, compared to the size of the primary spirals (see [10] for some qualitative explanations based on the multipole expansion of the potential). Thus, the potential can be presented in the multipole form

$$\Phi_1(r,\varphi) = A_\Phi(r)\cos 2(\varphi-\varphi_0) \simeq -r^n \cos 2(\varphi-\varphi_0)$$

not only at sufficiently large radii, but almost up to the primary spirals. Evidently, the mentioned correspondence between maxima of the response and the potential holds in the whole multipole region.

In the narrow transitional region between the spiral and the multipole regions (see Fig. 2) the response switches between minima and maxima of the potentials. The pitch angle of the spiral response is small due to the narrowness of the transitional region, and the angular length is apparently about $\pi/2$. Thus, the response in the transitional region is QTS.

QTS cannot contain any significant amount of old stars of the disk, since it contradicts to the observed small pitch angles of QTS. Indeed, QTS occur in the gas and young stars. The velocity dispersion of young stars increases with time and stars should simply leave the QTS region, otherwise the latter would be much wider.

### 2.2 Model Examples

In this section, several model examples will be considered. Figure 3 presents the response to the logarithmic spirals $\sigma_1 \propto e^{iB\log r}$ derived numerically using the exact formula (1) and the general expression for the simple layer potential.

The parameter $B$ defines the pitch angle $i$ of a spiral: $\tan i = 2/B$. In the first two figures, the primary spirals are open, while the last figure shows a tightly wound primary spiral. The responses in the figures follow the maxima of the

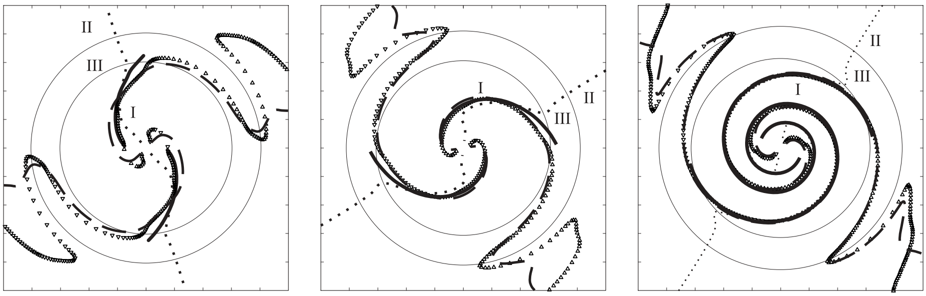



**Fig. 3.** The response of the galactic disk to the gravitational potential of the logarithmic spirals $\sigma_1 \propto e^{iB \log r}$. The spiral, multipole, and transitional regions are divided by the circles. Solid lines show the maxima of the primary spirals, dotted lines — minima of the gravitational potential, dashed lines — maxima of the second derivative of the gravitational potential, triangles — maxima of the response, calculated using (1).

second derivatives of the potential in accordance with the elementary theory. For open primary spirals, the QTS are evident. On the contrary, for the tightly wound primary spirals, it is more difficult to reveal QTS.

The pitch angle of the QTS seems to be roughly independent of the pitch angle of the primary spirals. For all these cases the pitch angles are about 10°.

## 3 Example of QTS in the Galaxy NGC 3631

NGC 3631 is rather bright non-barred galaxy of type Sc with well-defined spiral structure, observed nearly face-on.

As it was shown in [11], the second Fourier harmonic dominates over the others for this galaxy. It can be seen from Fig. 4a, where the contribution of the individual Fourier harmonics to the brightness deviation from axial symmetry is presented. Figure 4b shows the maximum of the $R$-map second Fourier harmonic superimposed on the image of the galaxy. The quarter turn spirals are clearly seen (this is also true for other bands — B, V, R, H$\alpha$). In the Fig. 4b the inner QTS are clearly seen. They arise just from the same mechanism as the outer QTS due to the decreasing of the primary spiral amplitude to the center of the disk.

The axisymmetric disk density profile and the primary spiral density can be inferred from the analysis of a brightness map, assuming the light to mass ratio to be constant. The obtained amplitude and phase of the second Fourier harmonic is given in Fig. 5. Thick curves show the smoothed functions that describe the primary spiral. The potential is restored by using the general formula of a simple layer. The rotation curve and the pattern velocity of the spiral structure is taken from [11], [12].

Figure 6 shows the response to the gravitational potential of the primary spirals for the grand design galaxy NGC 3631. It is seen that the response repeats the curves of the second derivative of the potential. The comparison of the

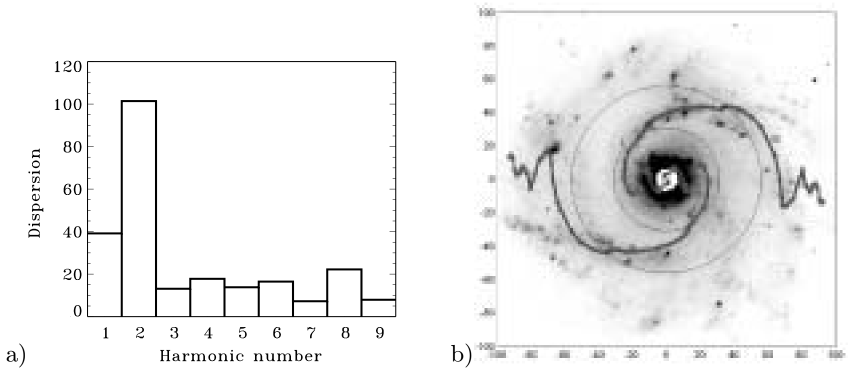



**Fig. 4.** a) Contributions of the individual Fourier harmonics to the brightness deviation from axial symmetry; b) maxima of the $m = 2$ Fourier harmonic superimposed on the $R$-map of the galaxy (ING archive). Circles divide the regions of inner QTS (I), spiral (II), and outer QTS (III).

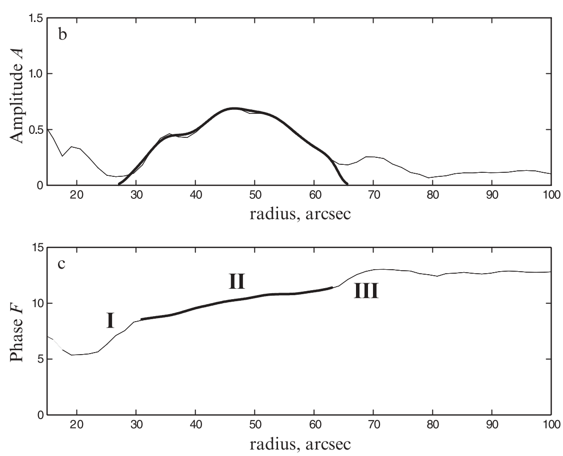



**Fig. 5.** The amplitude (in the arbitrary units) and phase (radians) of the $m = 2$ Fourier harmonic of the $R$-band image of the galaxy NGC 3631. Thick curves show the amplitude and phase of the primary spirals used in calculations.

response with the second Fourier harmonic (solid line) demonstrates the qualitative agreement of theory with observations. In the outer part, the response is somewhat longer, but the pattern structure of the QTS is reproduced. In the center, the response practically coincide with the Fourier harmonic: the length, the pitch angle and even the radially aligned structure in the very center is reproduced. Thus for the inner part, the good agreement is observed.

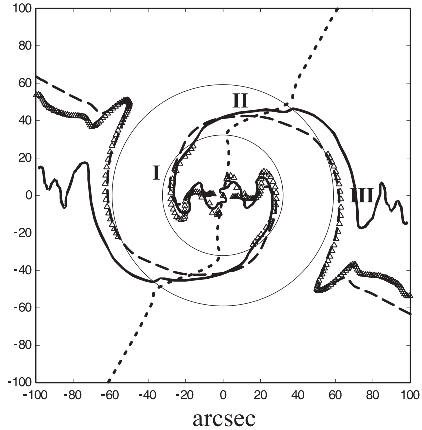



**Fig. 6.** The response of the grand design galaxy NGC 3631. Circles divide inner QTS, spiral, and outer QTS regions. Dotted lines show minima of the gravitational potential, dashed lines — maxima of the second derivative of the gravitational potential, triangles — maxima of the response. Solid lines show maxima of the $m = 2$ Fourier harmonics as in the Fig. 4b.

# 4 Conclusions

1. The phenomenon of QTS at the end of the primary arms of normal grand design galaxies has been analyzed. It is shown that:

- QTS are the universal nonresonance response of the galactic disk to the potential of the primary spirals;
- characteristic observed features of the QTS are explained by the peculiarities of the potential behavior near the end of the primary spirals.

2. QTS occurs not only in normal spiral galaxies. The example of the barred galaxy NGC 1365 is given in [10], [13]. When the bar is formed, the primary spiral must appear as the disk response to the potential of the bar. For the resonance excitation [15], the length of this spiral does not exceed $\pi/2$ for the standard fast bar, and $\pi$, for the Lynden-Bell slow bar [14], [15]. If the primary spiral is strong enough to change the multipole potential of the bar to the spiral-type potential, then the first quarter-turn spirals should form that elongate the primary spirals over $\pi/2$.

3. QTS provides a natural explanation for the sequential elongation of spirals both in normal and barred galaxies. The development of the QTS can make sufficiently powerful so that the total potential will turn into the spiral form once again. The secondary quarter-turn spirals will then appear, and the spiral structure elongates further over $\pi/2$.

**Acknowledgments**

This research has made use of the NASA/IPAC Extragalactic Database (NED) which is operated by the Jet Propulsion Laboratory, California Institute of Technology, under contract with the National Aeronautics and Space Administration and was partially based on data from the ING Archive. The work was supported in part by grants No. 99-02-18432 and 00-15-96528 provided by RFBR.

## References

1. D.L. Block, G. Bertin, A. Stockton, P. Grosbøl, A.F.M. Moorwood, R.F. Peletier: Astron. & Astrophys. **288**, 365 (1994)
2. P.J. Grosbøl, P.A. Patsis: Astron. & Astrophys. **336**, 840 (1998)
3. D.M. Elmegreen, F.R. Chromey, B.A. Bissell, K. Corrado: Astron. J. **118**, 2618 (1999)
4. A.M. Fridman, A.B. Mikhailovskii: JEPT **61**, 457 (1971)
5. G.S. Bisnovatii-Kogan, A.B. Mikhailovskii: Sov. Astron., **17**, 205 (1973)
6. C.C. Lin, F.H. Shu: Astrophys. J. **140**, 646 (1964)
7. A. Toomre: Astrophys. J. **139**, 1217 (1964)
8. G. Contopoulos, P. Grosbøl: Astron. & Astrophys. **155**, 11 (1986)
9. L.D. Landau, E.M. Lifshitz: *The classical theory of fields* (Pergamon Press, London, New York 1976), sect. 41
10. V.L. Polyachenko, E.V. Polyachenko: Astron. Rep., **46**, 1 (2002)
11. A.M. Fridman, O.V. Khoruzhii, E.V. Polyachenko, et al.: Mon. Not. R. astr. Soc. **323**, 651 (2001)
12. A.M. Fridman, O.V. Khoruzhii, E.V. Polyachenko, et al.: Phys. Letters, **A264**, 85 (1999)
13. E.V.Polyachenko: Mon. Not. R. astr. Soc. **331**, 394 (2002)
14. D. Lynden-Bell: Mon. Not. R. astr. Soc. **187**, 101 (1979)
15. I.I.Pasha, V.L. Polyachenko: Mon. Not. R. astr. Soc. **266**, 92 (1994)

# Dynamics of Galaxies: From Observations to Distribution Functions

Herwig Dejonghe and Veronique De Bruyne

Astronomical Observatory, Ghent University, Krijgslaan 281, S9, 9000 Gent, Belgium

**Abstract.** An overview is given of the currently most used dynamical modelling methods, with an emphasis on those methods that allow to derive a distribution function from observations of elliptical galaxies. Special attention is paid to the applications of distribution functions in the study of the internal dynamical structure of galaxies. It is indicated how existing modelling methods can be improved to correspond better with state-of-the-art observations and computation facilities.

## 1 Introduction

In general terms, a dynamical model for a stellar system provides a statistical description of a gravitational system, that is based as much as possible or practical on dynamical theory. More specifically, the best one can hope for is the determination of the distribution function (hereafter DF), which provides the probability to find a star with a given position and a given velocity. Such a model is supposed to provide a (good) approximation of the observed morphology and kinematics of a galaxy. In many cases, a dynamical model is only a vehicle to determine global parameters (e.g. total mass) or special characteristics (e.g. the presence of a central black hole). However, if dynamical modelling also involves the determination and interpretation of a DF, more detailed and structural information about the galaxy can be obtained. It is the purpose of this contribution to show that this goal is coming within reach.

## 2 General Considerations

### 2.1 Scope

Dynamical modelling always implies a simplification of some sort because in a realistic situation dynamical theory is virtually absent or impractical to apply. An obvious way out is the use of $N$-body simulations, which can righteously claim a generality and applicability that theory cannot match. Moreover, the sampling of a DF from an $N$-body simulation is rather trivial in principle. On the other hand, our capabilities to interpret huge amounts of data remains rather limited, though the amount of information one can handle keeps increasing. One should also keep in mind that (the possibility of) the presence of dark matter, the nature of which still eludes us, may vitiate many a, otherwise state-of-the-art, simulation, because it is by no means obvious if and how dark matter can

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 243–262, 2003.

be modelled by means of particles. We will not discuss the $N$-body technique in this contribution.

Many dynamical analyses restrict themselves to the study of the Jeans equations, which are integrals of the Liouville equation and thus are differential equations involving moments of the DF. The Jeans equations therefore provide the tool to study galaxies as hydrodynamical systems, filled with a "stellar gas". However, they do not generally form a closed set of equations, implying that they generally require a priori assumptions on the anisotropy of the dispersion tensor. They also cannot account for the fact that stars are on orbits, that link rather distant regions of the galaxy, by virtue of the fact that the Jeans equations are local because they are differential equations. As a consequence, more hydrodynamical models can be made than models based on DFs, not all of them however with a positive DF. Despite the extensive use that has been made of the Jeans equations in the past (e.g. [9], [46], [56]), we will not discuss them in this review.

### 2.2 Spiral Galaxies

As of now, relatively little has been done to determine DFs in spiral galaxies, at least compared to elliptical galaxies. There are many reasons for this. To begin with, there is generally a lot of extinction in spiral galaxies, which makes observations difficult. Spirals also tend to be less massive than ellipticals, and therefore high spectral resolution is needed to determine the relatively small velocity dispersions. The observational capabilities to accomplish this have only recently become practical. Moreover, there is the presence of the spiral arms. While it is well established that these are only perturbations on an otherwise rather smooth distribution, their presence, together with the rather patchy dust distribution associated with them, does not directly make the analysis any simpler. Last but not least, because of the spiral structure that must be rotating, spiral galaxies are not in a photometric dynamical equilibrium, and therefore the potential must be time dependent. This makes the modelling quite hard (though not impossible, see e.g. [78], [79]). Therefore, we will also limit our focus to elliptical galaxies. In the end, elliptical galaxies may prove to be equally complicated as spiral galaxies, but at least, from a photometric point of view, they look rather relaxed and smooth, and it does not seem therefore to be a gross simplification to assume that their underlying gravitational potential is time independent. Even if it is not (and surely it isn't), it is unlikely to evolve on timescales that are comparable to the crossing times of an individual orbit in an elliptical galaxy, as seems to be indicated by the presence of rather standard-looking ellipticals up to high redshift.

There is a rather extensive literature on the dynamical structure of our Galaxy. In many respects, it involves all issues that are relevant for spiral galaxies (e.g. the study of the central bar, the 3 kpc ring...), and many that are more typical for ellipticals (the structure of bulge and halo). It is clearly not within the scope of this review to cover all aspects of Galaxy research, and we will only mention attempts to determine the orbital structure of populations in our Galaxy for as far as equilibrium DFs are involved.

### 2.3 The Gravitational Potential

A dynamical theory implies somehow a model for the gravitational potential that is supposed to govern the motion of baryonic matter. Unfortunately, little is known about the exact form of the underlying gravitational potential of an elliptical galaxy. There are several philosophies for including a potential into the modelling.

One of the basic assumptions concerns the symmetries in the total mass distribution (and therefore the gravitational potential) of elliptical galaxies. It must be that the potential is somehow connected to the luminous matter distribution, which can be estimated from deprojecting the projected surface brightness. The luminous matter distribution can be given as a parameterized function (e.g. Nuker density, or [21], [61], [76]), or as a sum of basis densities (MGE [8],[36]) or completely numerically. Similarly, the potential can be a parameterized expression (e.g. power law models [37], [38]) or a sum of basis functions ([17],[57]), or a purely numerical function.

Triaxial systems provide the most general description for elliptical galaxies. Due to their complexity they are not widely used to fit observations (but see [57], [58]), and their characteristics are more often studied on theoretical grounds (e.g. [28], [51], [60], [74]). We will not discuss this case any further.

Since the symmetries in the potential largely dictate the modelling flexibility that dynamical theory will allow, the sequel of this review is structured accordingly.

## 3 Spherical Potentials

In the spherical world view, the simplest models for elliptical galaxies have DFs that depend only on the energy $E$ and therefore have an isotropic dispersion tensor. For the construction of this type of models, it suffices to have an expression for the spherical spatial density (and hence the potential through Poisson's equation), and to solve an integral equations (Eddington's formula). There is a close connection between such models and spherical isotropic models based on the Jeans equations. In spite of the fact that the DF is a more fundamental quantity than the velocity dispersion, this stellar dynamical description of elliptical galaxies, in the spherical and isotropic framework at least, and for as far as it is a derivative of observational information, has failed to make important contributions in addition to what has been obtained with the Jeans equations.

Spherical models can be made anisotropic by introducing the modulus $L$ of the angular momentum in the DF $F(E, L)$. There are many anisotropic models possible that reproduce the same mass density and velocity dispersion ([25], [26]). While the anisotropic generalization is fairly obvious from a mathematical point of view, it is quite unclear why a stellar system would form and settle in a form that is completely blind for the orientation of the orbital planes of its orbits (orbits in spherical potentials are planar). Nevertheless, such models learn us that there is a rather bewildering variety of dynamical models that could fit the photometry in a spherical geometry, though the observed velocity dispersion could be quite different. It remains a mystery why nature chooses to realize only a subset of them, so as to make the fundamental plane thin.

Finally, the presence of dark matter offers the possibility to consider non-spherical models within spherical potentials, the idea being that any inconsistency will be taken care of by the dark matter. It is clearly hard to defend this assumption in any rigorous way, but it is certainly not improbable that the dark matter is rounder than the luminous matter. Such models could therefore be seen as a useful approximation of the real thing [20], [58].¡

# 4 Axisymmetric Potentials

## 4.1 Two-Integral Dynamical Models

In the axisymmetric paradigm, the component of the angular momentum parallel to the symmetry axis $L_z$ is in general the only conserved integral besides the energy. Two-integral models with a DF of the form $F(E, L_z)$ have equal radial and vertical velocity dispersion, but these are different from the tangential dispersion. Such DFs consist of an even part in $L_z$ and and odd part in $L_z$. The spatial density does only depend on the even part of the DF. As a consequence, if a DF generates a system with given spatial density, an unlimited number of DFs can be constructed that generate that spatial density by adding odd functions in $L_z$ to the DF. The odd part of the DF is important for the rotation in the galaxy.

The analytical theory concerning the determination of the DF from the mass density and the mean rotation has been initiated in [49], [25]. Not many of these analytical techniques allow to include the observed dynamics in an easy way, and therefore they have not been applied very often. A contour integral formula for the calculation of two-integral DFs for axisymmetric systems, derived directly from the density has been developed by Hunter & Qian [50]. The method requires an analytical expression for the potential for the density in terms of the potential. The DF can be calculated as a contour integral. Applications can be found in e.g. [63], [82].

Kuijken ([54]) developed a completely numerical technique for the construction of DFs. The numerical inversion of the integral equations connecting the spatial density and the streaming velocity with the DF, requires some smoothing which is achieved by a parametrization of the DF in continuous bilinear segments in this case.

When inhomogeneous data sets containing photometric and kinematic informations are considered, a quadratic programming algorithm [27] is a practical tool to obtain a dynamical model. For applications, see [35], [68], [69].

Two-integral DFs can also be constructed using the Richardson-Lucy algorithm ([22]). Furthermore, there are also a number of papers that describe classes of analytical models, but that are not really used for modelling data so far ([10], [11], [12], [52], [65]).

A dynamical system with two integrals of motion does not provide enough freedom for the modelling of our galaxy: observations have shown that in the solar neighbourhood the radial and vertical velocity dispersion are clearly different; hence the need for an additional degree of freedom. Moreover, numerical integration of orbits in axisymmetric systems has revealed that most orbits obey a third integral.

The introduction of such a third integral into dynamical models makes the modelling considerable more complicated, mainly because a third integral can only be determined approximately. There are several ways to approximate a third integral of motion. Some of these techniques use perturbation methods, e.g. [23], [44], [59] or derive a so-called 'partial integral', e.g. [34], [39]. Another option is the use of Stäckel potentials, where the expression for a third integral is analytically known. Some commonly used techniques in three-integral dynamical modelling are mentioned here.

### 4.2 Three-Integral Dynamical Models

**Schwarzschild Methods.** Schwarzschild [67] proposed a method that relies on the numerical calculation of a large library of orbits in a given potential. During integration, the characteristics of the orbits (e.g. spatial and projected density, line profiles, velocity moments) are stored on a grid. The weighted orbits are then combined in order to approximate the observed quantities as good as possible. The orbits are parameterized by their starting points, that are connected to integrals of motion. Hence, there is no need for an expression for an effective third integral.

This method is very general because there are no prerequisites for the expression for the potential or the DF. On the other hand it is a computationally demanding method and it is not straightforward to obtain a sufficiently smooth DF. In most applications the main issue is to reproduce the velocity moments and line profiles in order to estimate a (central) mass for the galaxy. Applications of this method can be found in [7], [8], [13], [14], [15], [41], [42], [64], [76], [81].

Whereas the method is independent of a priori assumptions, it does require a substantial degree of regularization. One way to make this easier is to use building blocks that are smoother than individual orbits [60]. One step further is to use two integral DF components whose observables can be calculated as integrals of the DF without going through the step of orbit integration, but also here regularization is required in order to derive smooth DFs [80]. Models that are created in this way give up a great deal of generality of the Schwarzschild method and are closely related to another widely used type of models, those that consist of a combination of basis function DFs.

**Basis Function DFs.** Independently of how the third integral of motion is approximated, DFs can be created as a sum of basis functions. This technique relies on the linearity of the equations and on a minimization algorithm to find the best fitting DF (e.g. [16], [18], [23], [24], [30], [31], [36], [45], [53], [55], [58], [59], [66], [71]).

Within this class of dynamical models a range of different types of basis functions can be used. Depending on the specific applications, the number of assumptions may be modest. Expansion methods offer the possibility to obtain non-parametric fits. The most important advantage however is that they can deliver smooth and positive DFs that are explicitly known, thereby offering possibilities that are hard to retrieve from models that rely on pure numerical integration of orbits. These properties make this kind of modelling rather appealing if the goal is to get insight in the internal structure of elliptical galaxies (e.g. [18]).

**The Use of Separable Potentials.** Potentials for which the Hamilton-Jacobi equations are separable in a given orthogonal coordinate system are called separable Stäckel potentials. For this family of potentials, three integrals of motion are explicitly known for all initial conditions, and the expressions for them are analytical and simple. Moreover, the orbital families in Stäckel potentials are easily classifiable and correspond to orbits that are found in more general potentials. In many cases they provide a good general description of elliptical galaxies. Despite the fact that these are not the most general potentials, studying models based on them may provide valuable insight in the internal structure of elliptical galaxies ([32], [29], [30]). However, there are no irregular orbits in these potentials [33]. The explicit knowledge of a third integral of motion allows to write down the moments of judiciously chosen DFs in an analytical way. This is very practical in combination with the use of basis function DFs (e.g. [18], [30], [36], [62], [71]).

## 5 What Can We Learn from Distribution Functions?

In most of the cases where three-integral DFs are explicitly shown (e.g. [8], [30], [59], [81]) this is done to illustrate that there is indeed structure in the third integral and to draw conclusions concerning the degree of anisotropy of the stellar system. However, not all of the presented DFs are smooth enough to get more detailed information on the internal dynamical structure of the galaxy.

Figure 1 gives an illustration of how information in a three-integral DF can be visualized, in this case by means of intersections with a plane of constant third integral in integral space or in turning point space. A representation in turning point space seems to be the most intuitive, since it can be most easily linked to physical orbits and their spatial extent. However, when a modelling procedure is adopted that makes explicitly use of the integrals of motion, a representation in integral space gives more insight in how the model has used the freedom available in the basis DFs. Both representations are complementary, in the sense that radial orbits are highlighted in a representation in the $(E, L_z)$-plane, while a representation in turning point space gives a clearer view on circular orbits.

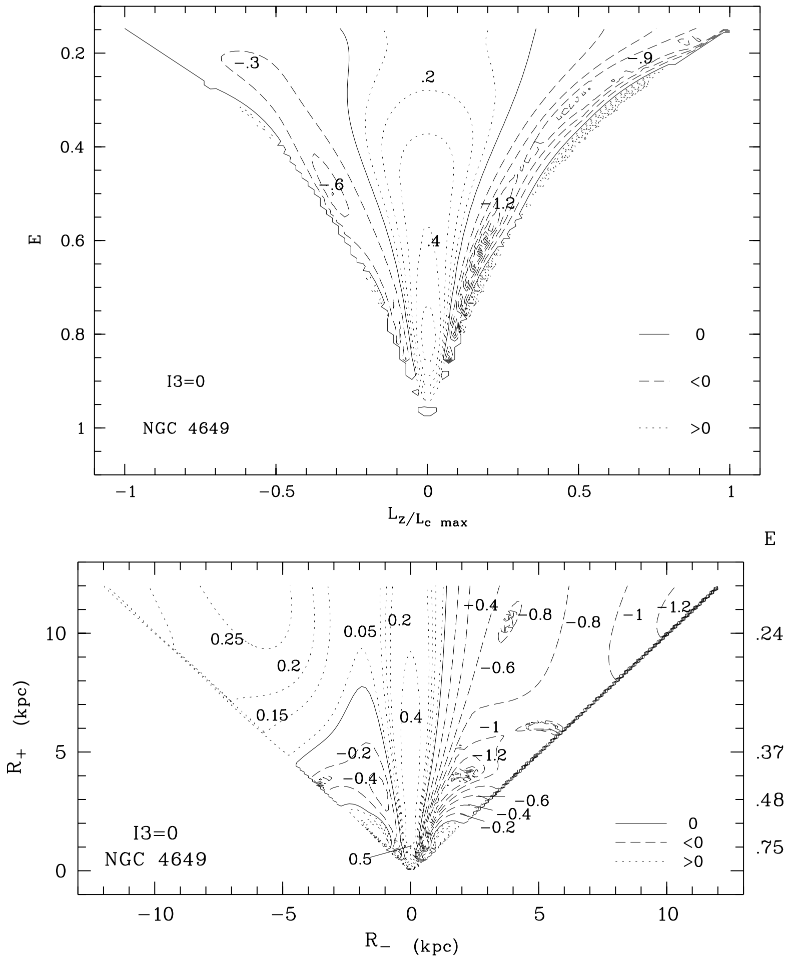



**Fig. 1.** An example of visualization of (normalized) DFs. Upper panel: representation in integral space of an intersection of a three-integral DF with a plane of constant third integral ($I_3 = 0$). Lower panel: representation of the same quantity in turning point space. Here, $R_+$ is the apocentre in the equatorial plane, $R_-$ is the pericentre in the equatorial plane, the sign of $R_-$ is equal to the sign of $L_z$.

### 5.1 Disentangle Photometric and Kinematic Information

Since a dynamical model is based on a mix of photometric and kinematic observables, the resulting DF carries information on both. When dynamical models are calculated for elliptical galaxies that behave kinematically different, the DFs should also be significantly different. When these galaxies have different photometric properties, a key question is whether there are significant differences between the DFs that can be attributed to the kinematic information alone. To answer such a question, it is necessary to eliminate from the DFs the signature of the photometry.

Such a normalization can be realized by dividing the three-integral DF, obtained by fitting the photometry and kinematics, by the even two-integral DF that is determined completely by fitting the photometry only. In addition, such a normalized DF is dimensionless and coordinate independent, which makes it a suitable quantity for comparison between different dynamical models.

Figure 1 shows a logarithmic representation of such a normalized three-integral DF for NGC 4649. A detailed description of the modelling and the observations can be found in [18]. The main conclusion from the figure is that orbits with small $|L_z|$ seem to play an important role in this model (largest values for the contours in a vertically stretched region in the middle of the contour plot), and this is an effect attributable to the kinematical information only. In the representation in turning point space, a small region with positive contours for circular or near-circular orbits with negative $L_z$ becomes clear, although the observed mean velocity is positive. This shows that small amounts of counterrotation in elliptical galaxies may not leave their imprint on the observed kinematics.

### 5.2 Additional Classification Parameter

The class of elliptical galaxies is generally seen as a twodimensional manifold, characterized by flattening and boxyness/diskyness, which are photometric parameters. One may think of refining the classification by using kinematic information. This could be done by means of comparing normalized DFs as described in the previous part. In order to quantify this idea, a diagnostic could be used:

$$\nu = \frac{log(M_1) - log(M_2)}{2\sqrt{\sigma^2_{M_1} + \sigma^2_{M_1}}}, \tag{1}$$

with $\log(M_i)$ the logarithm of the normalized DF for model $i$ and $\sigma_{M_i}$ the error on the normalized DF for model $i$.

An illustration of what can be concluded from this diagnostic can be found in Fig. 2, where a comparison between normalized DFs of NGC 4649 and NGC 7097 indicates that it is indeed possible to distinguish between galaxies based on their kinematics alone. The figure shows that the orbital density of radial orbits and circular orbits is significantly high for NGC 4649, while orbits with moderate to high $|L_z|$ are preferred in the model for NGC 7097. These differences in orbital densities are due only to differences in kinematic parameters.

### 5.3 Counterrotation in NGC 7097

NGC 7097 is a disky elliptical where the mean velocity in the inner 1.76 kpc indicates a counterrotating core (lowest $\langle v \rangle$ is -20 km/s at 0.88 kpc). This galaxy was modeled with a three-integral model based on a Stäckel potential [18].

A representation in turning point space of the DF where the number of orbits with $L_z < 0$ is subtracted from the number of orbits with $L_z > 0$ (see left panel

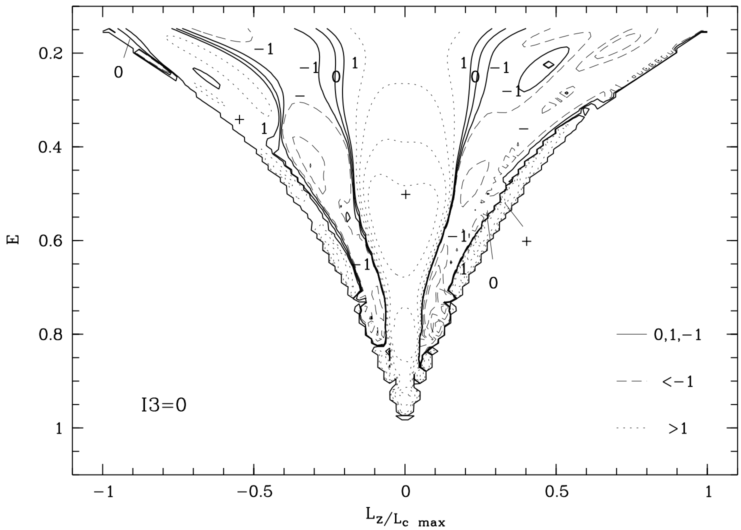



**Fig. 2.** Comparison of normalized DFs in a plane with zero third integral, by means of the diagnostic $\nu$, as defined in equation 1. This figure compares DFs for NGC 4649 and NGC 7097. Negative contours are in dashed lines and indicate that ($DF_{\mathrm{NGC\ 7097}} > DF_{\mathrm{NGC\ 4649}}$), positive contours are in dotted lines and indicate that ($DF_{\mathrm{NGC\ 4649}} > DF_{\mathrm{NGC\ 7097}}$). Solid lines display contours for -1,0,1.

in Fig. 3) revealed that the counterrotation in this galaxy is not necessarily caused by a compact group of stars. Stars contributing to the counterrotation can be found as far as 4 kpc from the centre, while the mean velocity profile shows only counterrotation up to 1.76 kpc. For a toy galaxy where the amount of rotation and counterrotation is doubled with respect to the counterrotation present in NGC 7097, the counterrotating orbits are confined to a much smaller spatial region (right panel in Fig. 3). This indicates that the amount of signal in the $\langle v \rangle$-profile for NGC 7097 is not enough to constrain the counterrotating stars to the central part of the galaxy. This has also implications for possible formation scenarios for ellipticals with counterrotating cores.

## 6 Continuous Improving on Modelling Techniques

Dynamical modelling is a vivid research field, where considerable efforts are spent on improving existing modelling methods. These continuous developments are triggered by practical considerations such as ever improving computational capabilities and/or observational facilities, that deliver data of ever increasing quality and quantity.

### 6.1 Spectra as Source of Information

The shape of the absorption lines in a galaxy spectrum depends on the composition of the stellar mix and the dynamics of the galaxy. The most frequently used

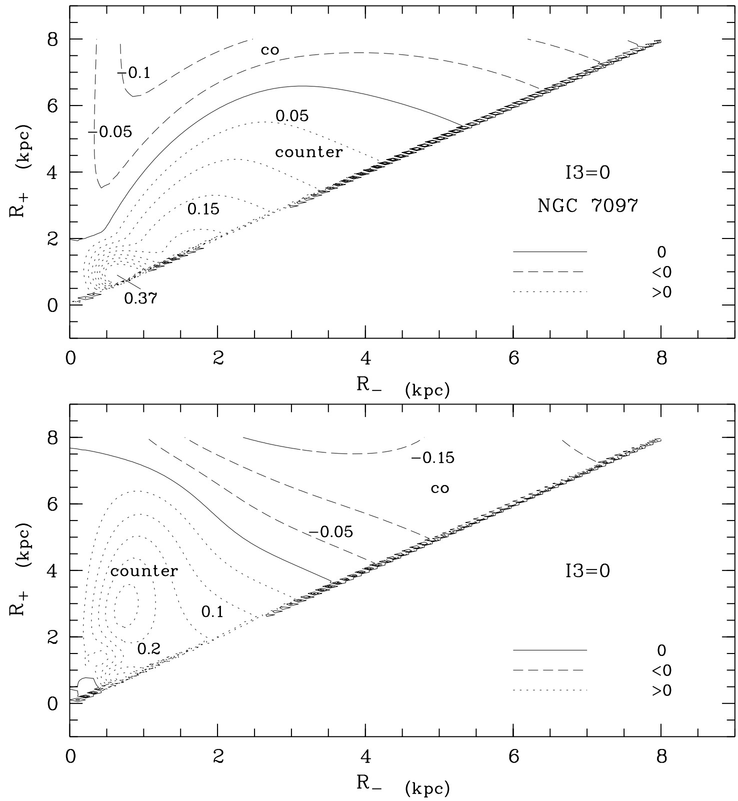



**Fig. 3.** Counterrotation in NGC 7097. Normalized DFs in turning points space, where the number of orbits with negative angular momentum is subtracted from the number of orbits with positive angular momentum. For NGC 7097 in the left panel: this galaxy contains counterrotating orbits with radii up to 4 kpc. The right panel shows the case for a toy galaxy, where the amount of rotations was doubled with respect to the case of NGC 7097.

approach to dynamical modelling is a two-step process: first the line-of-sight velocity distribution (hereafter LOSVD) is determined from the observations and kinematic parameters are derived. In a second step these parameters are used to constrain a DF.

A new strategy uses the spectra as they come to constrain (1) a DF and (2) a template mix in one modelling process [31]. This is equivalent to the construction of a detailed dynamical model and a population synthesis at the same time. However, fitting directly to spectra implies a considerable increase in the number of data points. It seems that only when analytically tractable basis functions are used to construct the DF, in which case the contribution of the dynamical components to the spectra can be calculated analytically, this is a feasible setup.

With this method, a three-integral dynamical model has been constructed for NGC 3258 [20]. Spectral features from the Ca II triplet were modeled using

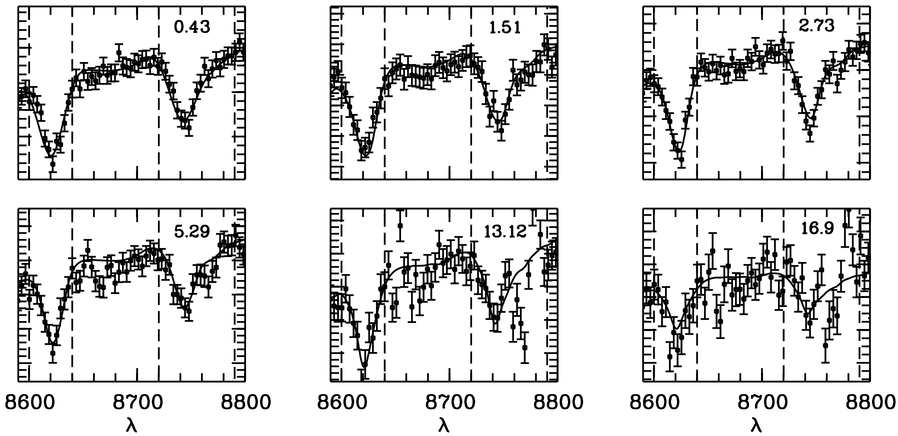



**Fig. 4.** Dynamical model for NGC 3258 fitted directly to spectral features from the Ca II triplet (data points with error bars) and the fit (in solid lines). The fit regions are located between the intervals indicated by dashed lines with centre around 8620 Å and 8750 Å. Projected radii are indicated in the top right corners.

two different template stars (a G5 dwarf and a K4 giant). Figure 4 shows the fit to the two strongest lines of the Ca II triplet at different projected radii.

Figure 5 shows some results of the modelling: the number density of the different populations (upper panel) and the kinematic parameters (projected and spatial) for the different populations. The obtained DF can be used to study the dynamics of the separate stellar populations in the galaxy. This can be done by comparing their spatial kinematics (see Fig. 5), their LOSVDs (see Fig. 6) or representations of their DFs (see Fig. 7).

The plot with relative densities in Fig. 5 shows that the centre contains mainly K4 giants, between 0.5 and 2 kpc there is almost an equal amount of both stellar types and for larger radii the K4 giants are again dominating the stellar light distribution. As for the kinematics, it is clear that the G5 dwarfs rotate more than the K4 giants do (illustrated in Figs. 5 and 6). The anisotropy parameter shows that the model is isotropic in the centre, becomes radial anisotropic soon and becomes tangential anisotropic at 2 kpc.

The shapes of the line profiles for the total DF (solid lines in Fig. 6) are not Gaussian but have a large variety of shapes. These shapes can be better understood when the LOSVDs for the total DF and the ones for the separate populations are compared. The LOSVDs for the total DF are composed of the LOSVDs of the G5 dwarfs (dotted profiles) and the K4 giants (dashed profiles). The relative heights of these profiles indicate the relative number densities of the stellar populations.

Intersections for the DFs with the equatorial plane are shown in Fig. 7, presented in the $(E, L_z)$-plane and in turning point space. Also here, the characteristics of the total DF (upper row) are a mix of the characteristics of the DFs for the separate stellar populations.

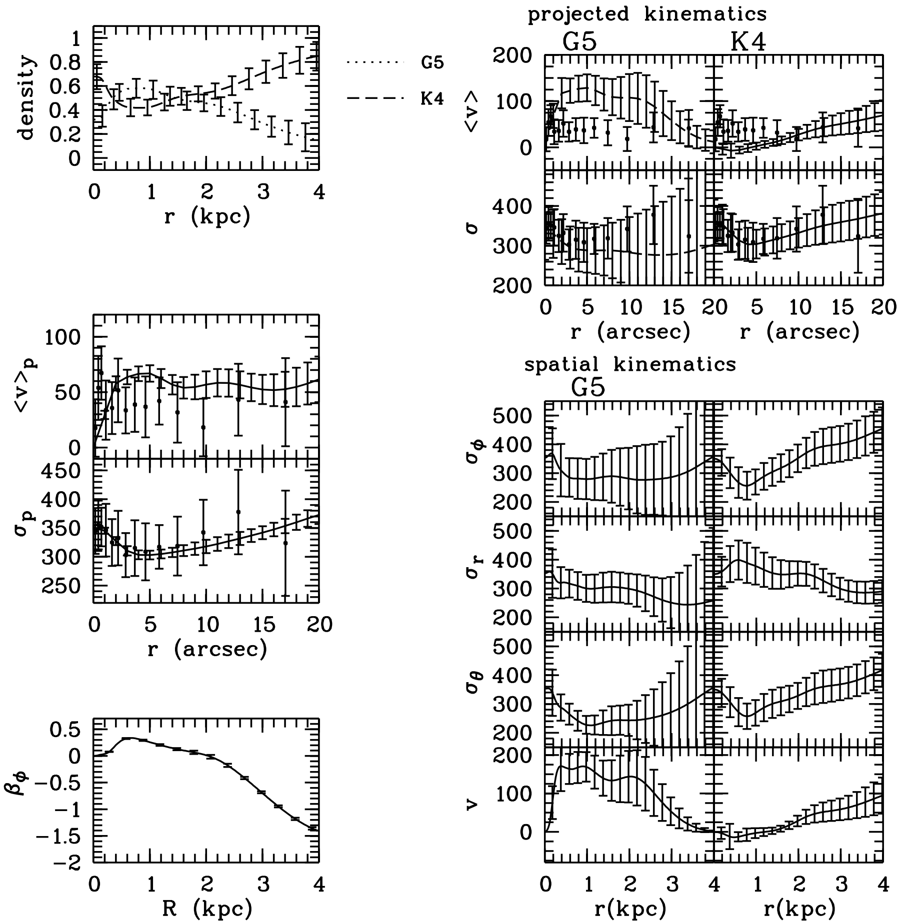



**Fig. 5.** Dynamical model for NGC 3258. Upper left panel: densities of the stellar populations. Middle left panels: projected mean velocity and projected velocity dispersion for the total DF in solid line and with error bars. The data points presented by dots are kinematic parameters obtained from fitting a Gauss-Hermite series up to fourth order to the observed spectra. Lower left panel: The anisotropy parameter for the model. Upper right panels: Projected kinematics with error bars for the stellar templates (G5 dwarfs on the left, K4 giants on the right), from top to bottom: projected mean velocity, projected velocity dispersion. The data points presented by dots are kinematic parameters obtained from fitting a Gauss-Hermite series up to fourth order to the observed spectra. Lower right panels: Spatial kinematics with error bars for the stellar templates (G5 dwarfs on the left, K4 giants on the right), from top to bottom: spatial dispersions and spatial velocity.

### 6.2 Construction of a Third Integral

Most dynamical models with DFs currently found in the literature use Stäckel potentials and the explicit expression for the third integral offered by these sys-

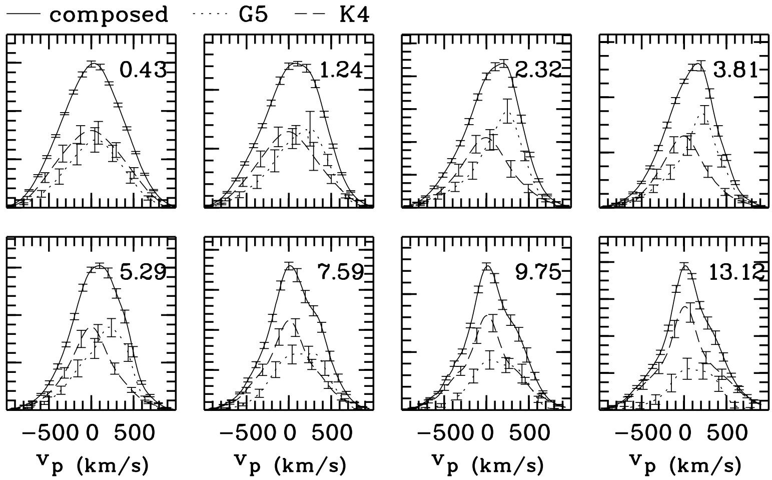



**Fig. 6.** LOSVDs and error bars calculated from the dynamical model for NGC 3258, for the total DF in solid lines, for the G5 dwarfs in dotted lines and for the K4 giants in dashed lines.

tems. To improve on the generality of this approach, one may think of approximating a galaxy potential by a set of Stäckel potentials instead of one single Stäckel potential.

This can be done in an elegant way, since creating a grid in energy and angular momentum in integral space creates a sequence of surfaces in space so that for fixed angular momentum, orbits with higher energy fill a volume that is completely embedded in the volume filled by an orbit with lower energy. Similarly, for fixed energy, orbits with larger angular momentum are completely within the volume filled by orbits with smaller angular momentum. Hence, dividing the $(E, L_z)$-plane into a number of rectangles is equivalent to dividing space into a number of bounded domains, see Fig. 8. For each of these domains, a Stäckel potential can be determined that is locally a good approximation to the original potential. The validity of the approximation can be checked in several ways, based on numerical integration of orbits in the original potential and fitted Stäckel potential. Various criteria to judge the agreement between the orbits in the original potential and local Stäckel potential are surfaces of section, conservation of $I_3$, orbital densities, topology of orbits. An application of this procedure can be found in [17].

## 7 Some Concerns Before Modelling Takes of

### 7.1 Influence of Dust

In the context of dynamical modelling, elliptical galaxies are traditionally considered as dust free stellar systems, but it is well known that in nature they do

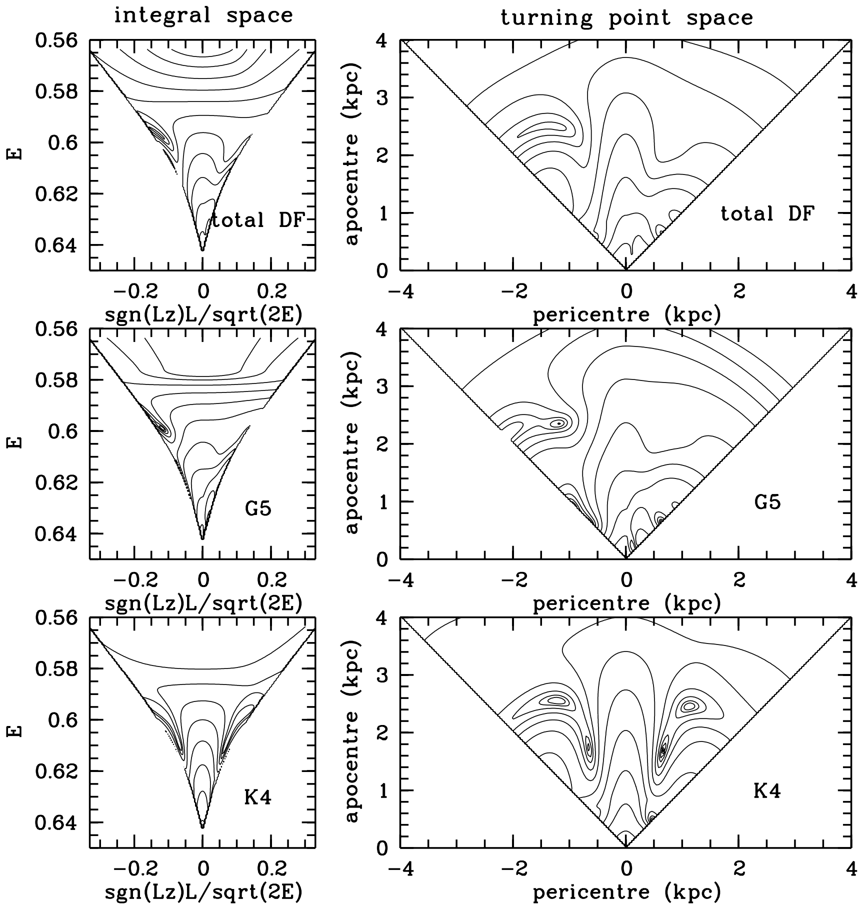



**Fig. 7.** Representations in integral space (left) and turning point space (right) of the DF for NGC 3258 in the equatorial plane. Here, $R_+$ is the apocentre in the equatorial plane, $R_-$ is the pericentre in the equatorial plane, the sign of $R_-$ is equal to the sign of $L_z$. Contours are equally spaced in log(phase space density). From top to bottom: for the total DF, for the G5 dwarfs, for the K4 giants.

contain various amounts of dust. Moreover, observed photometry and kinematics can be seriously effected by the dust.

There are numerous observations that prove the existence of dust in elliptical galaxies. Discrete optical dust features are seen in a number of ellipticals as dust lanes ([6], [48]) or in central regions ([40], [72], [77], [73]). Thermal far-infrared emission has revealed, unexpectedly, large IRAS fluxes and a comparison with optical dust features indicated that most of the dust has to exist as a diffuse component ([47]).

The dust grains present in ellipticals can cause absorption and scattering of the light. The result is that photons disappear from the line of sight and reappear on other lines of sight. The effect of this on the observed photometry and

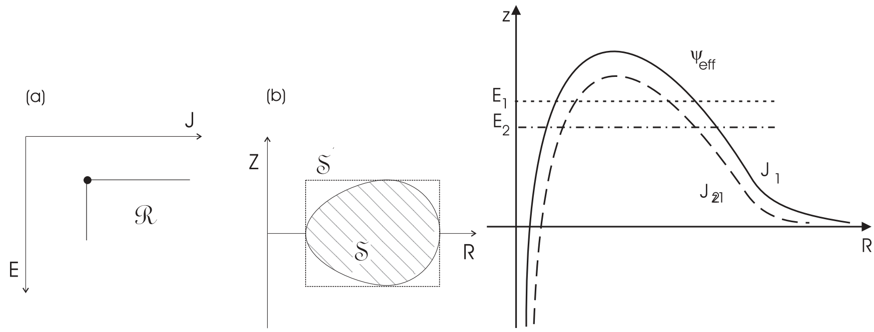



**Fig. 8.** Left: A grid in integral space (a) is a subdivision in space (b). Right: Intersections of lines of constant energy with the curve for the effective potential (solid line) illustrate that orbits with higher energy ($E_1$) fill a volume that is completely embedded in the volume filled by an orbit with lower energy ($E_2$). Similarly, for fixed energy (e.g. $E_1$), orbits with larger angular momentum ($J_1$)are completely within the volume filled by orbits with smaller angular momentum ($J_2$).

kinematics can be studied by solving the radiative transfer equation, taking dust attenuation into account and calculating the light profile and projected kinematics numerically ([1], [3], [4]). In practice however, a Monte Carlo routine that generates millions of photons turns out to be most cost-effective and flexible. It is able to simulate quite naturally the scattering process and it includes velocity information in an elegant way. Moreover, there are no restrictions on the geometry of stars and dust and a decent error analysis is possible.

This modelling showed that absorption and scattering by dust grains affect all observables. As for the kinematic profiles, there are no large effects in the central regions of the galaxy. On the other hand, there are serious effects on the kinematics in the outer regions: the projected velocity dispersion decreases more slowly while the $h_4$ profile increases quite dramatically (see Fig. 9). These effects are identical to what is generally considered as the kinematic signature of a dark matter halo. Hence, dust attenuation may reduce or even eliminate the need for a dark matter halo, giving rise to a new mass-dust degeneracy. This is investigated in more detail in [2]. Possible ways to break this degeneracy is to include dust attenuation in dynamical modelling and/or to use near-infrared kinematics which are less influenced by dust attenuation.

### 7.2 Error Bars on Data

Undoubtly, it is important to have photometry and kinematic data of good quality, but it is of equal importance to have a reliable error estimate on kinematic parameters. For dynamical modelling, these estimates are often used in goodness of fit indicators, and the decision whether to accept or reject a particular dynamical model often depends critically on the error bars of the data (e.g. [13], [64], [53]).

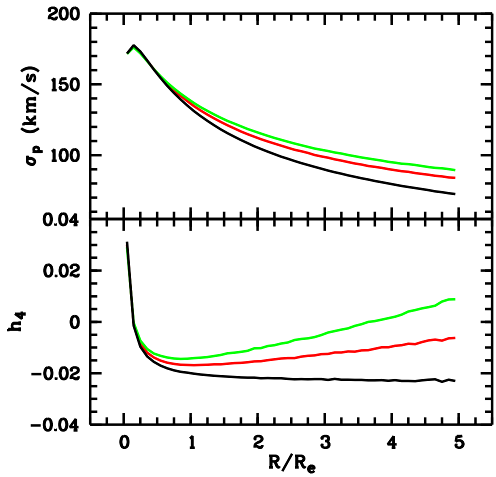



**Fig. 9.** Effect of dust attenuation on the projected kinematics, in particular the velocity dispersion and the $h_4$ profiles. The profiles are shown for a model without dust attenuation and for models with optical depths of V = 0.5 and V = 1.

An observed galaxy spectrum is the sum of individual stellar spectra of stars moving along the line of sight that have been redshifted by the rotation of the galaxy and smeared out by the velocity dispersion of the stars. Since the seventies, much effort has been spent on retrieving the LOSVD from observations [19]. In this context, it is convenient to express the information in that LOSVD in a limited number of parameters.

The higher order moments of the LOSVD depend critically on the wings of the profiles. Instead of expressing the characteristics of a LOSVD through its moments, a parameterized version of the LOSVD is often used in the form of a truncated Gauss-Hermite series ([43], [75]).

These parameters and an error estimate on them are often obtained using a least-squares minimization. The statistical interpretation of this method relies on the assumption that the noise is independent and Gaussian distributed on the input data. These conditions are generally not met after standard data reduction steps. As a consequence, the errors derived from standard statistics will in many cases differ from the real errors on the kinematic parameters.

Sometimes, Monte Carlo simulations are used ([5], [70]) to estimate the uncertainties on the derived parameters. For the realization of synthetic galaxy spectra, a Gaussian noise distribution with given S/N is used. So also this approach will show a similar tendency to underestimate the errors. People are aware of this, but it is mostly left unclear how large the differences between the estimated and real error bars are.

Recently, a method was proposed that makes a diagnosis of the characteristics of the real noise on the spectra and that allows to calculate more realistic error bars on kinematic parameters [19]. An illustration is given in Fig. 10. Panel (a) shows a galaxy spectrum and a fit to this spectrum, panel (b) shows the fitting

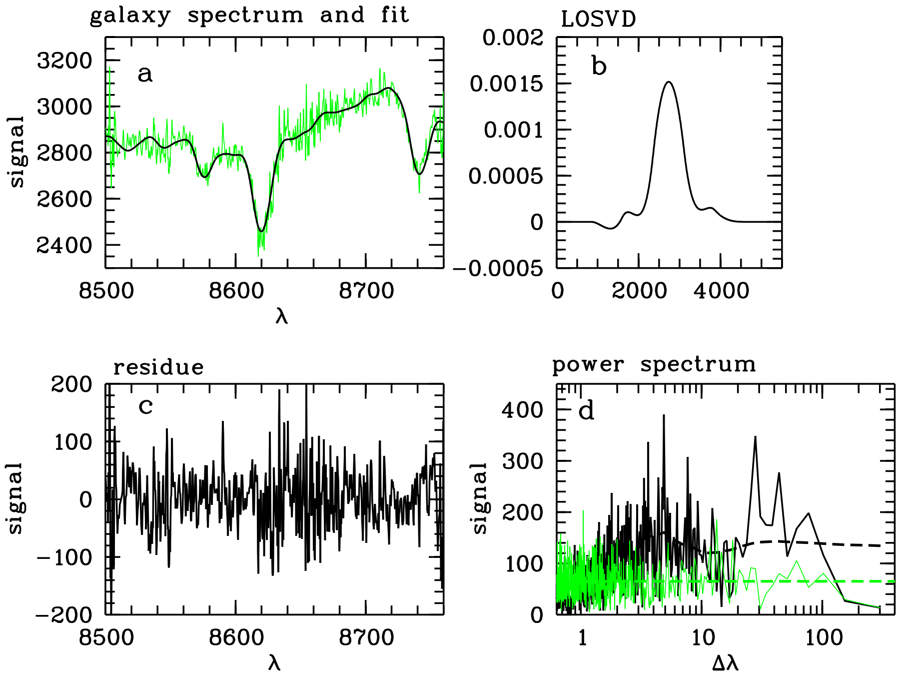



**Fig. 10.** Illustration of a technique that allows to calculate realistic error estimates. Panel (a): observed spectrum and fit. Panel (b): fitting LOSVD. Panel (c): residue of the fit. Panel (d): power spectrum of residue (black) and power spectrum with S/N of input spectrum (grey).

LOSVD and panel (c) shows the residual of the fit, that is considered as the real noise involved in the problem. Panel (d) shows the power spectrum of the residual in black and the power spectrum of a Poisson noise profile according to the S/N of the spectrum in panel (a) in grey. It is clear that the real noise involved is not Poissonian. The characteristics of the noise can be represented by the smoothed representation (dashed line). Realistic error estimates are obtained from Monte Carlo simulations with synthetic galactic spectra where the noise distribution has the same power spectrum as the real noise.

In the test case of NGC 3258, the realistic errors appeared to be almost a factor of 2 larger than the errors based on least squares statistics. Moreover, for the first time is was shown that the way the spectra are sampled in the data reduction has a non-negligible influence on the quality of the derived kinematics.

## 8 Conclusions

In this contribution we have given an overview of the most frequently used dynamical modelling methods. If detailed and structural information about a galaxy is to be obtained, the dynamical modelling should involve the determination of a DF. By means of a number of case studies it is shown how structural information contained in DFs can be visualized and interpreted. It is clear that this requires the DF to be a smooth function. However, not all dynamical modelling methods that are currently used deliver a DF that can be used for this purpose. If one agrees to use analytical approximations to the third integral (axisymmetric case), an expansion of the DF in analytic basis functions seems to offer only advantages.

We finally discuss a few recent developments that have become possible thanks to improved observational and computational capabilities, and we also critically reexamine some established paradigms. In particular, new astrophysical issues can be addressed if a direct fit to the spectra can be constructed. There are now possibilities to perform dynamical modelling and population synthesis in one go. We also explore what would need to be revised if one of the basic premises of dynamical modelling of elliptical galaxies is wrong: that there is little or no dust in elliptical galaxies. It turns out that the presence of dust has a non-negligible influence on the observable kinematics, especially in the outer regions that are crucial in our assessment of the presence of dark matter. Finally, we revisit the usual error analyses on the kinematic parameters, and show that in general, error bars of higher order kinematic parameters are underestimated by almost a factor of 2.

## References

1. M. Baes, H. Dejonghe, MNRAS, **313**, 153 (2000)
2. M. Baes, H. Dejonghe, ApJ, **563**, L19 (2001)
3. M. Baes, H. Dejonghe, MNRAS, **318**, 798 (2000)
4. M. Baes, H. Dejonghe, MNRAS, **335**, 441 (2002)
5. R. Bender, R.P. Saglia, O.E. Gerhard, MNRAS, **269**, 785 (1994)
6. F. Bertola, G. Galetta, ApJ, **226**, L15 (1978)
7. G.A. Bower, R.F. Green, R. Bender, K. Gebhardt, T.R. Lauer, J. Magorrian, D.O. Richstone, A. Danks, T. Gull, J. Hutchings, C. Joseph, M.E. Kaiser, D. Weistrop, B. Woodgate, C. Nelson, E.M. Malumuth, ApJ, **550**, 75 (2001)
8. M. Cappellari, E.K. Verolme, R.P. van der Marel, G.A. Verdoes Kleijn, G.D. Illingworth, M. Franx, C.M. Carollo, P.T. de Zeeuw, ApJ, **578**, 787 (2002)
9. P. Cinzano, R.P. van der Marel, MNRAS, **270**, 325 (1994)
10. L. Ciotti, ApJ, **520**, 574 (1999)
11. L. Ciotti, ApJ, **471**, 68 (1996)
12. L. Ciotti, B. Lanzoni, A&A, **321**, 724 (1997)
13. N. Cretton, F. C. van den Bosch F., ApJ, **514**, 704 (1999)
14. N. Cretton, P.T. de Zeeuw , R.P. van der Marel, H.-W. Rix, ApJS, **124**, 383 (1999)
15. N. Cretton, H.-W. Rix, P.T. de Zeeuw, ApJ, **536**, 319 (2000)
16. J. de Brijne, R. P. van der Marel, P. T. de Zeeuw, MNRAS, **282**, 909 (1996)
17. V. De Bruyne, F. Leeuwin, H. Dejonghe, MNRAS, **311**, 297 (2000)
18. V. De Bruyne, H. Dejonghe, A. Pizzella, M. Bernardi, W.W. Zeilinger, ApJ, **546**, 903 (2001)
19. V. De Bruyne, P. Vauterin, S. De Rijcke, H. Dejonghe, MNRAS, in publication
20. V. De Bruyne, S. De Rijcke, H. Dejonghe, W.W. Zeilinger, submitted to MNRAS

21. W. Dehnen, MNRAS, **265**, 250 (1993)
22. W. Dehnen, MNRAS, **274**, 919 (1995)
23. W. Dehnen, O.E. Gerhard, MNRAS, **261**, 311 (1993)
24. W. Dehnen, O.E. Gerhard, MNRAS, **268**, 1019 (1994)
25. H. Dejonghe, Phys. Rep, **133**, 217 (1986)
26. H. Dejonghe, MNRAS, **224**, 13 (1987)
27. H. Dejonghe, ApJ, **343**, 113 (1989)
28. H. Dejonghe, D. Laurent, MNRAS, **252**, 606 (1991)
29. H. Dejonghe, P.T. de Zeeuw, ApJ, **333**, 90 (1988)
30. H. Dejonghe, V. De Bruyne, P. Vauterin, W.W. Zeilinger, A&A, **306**, 363 (1996)
31. S. De Rijcke, H. Dejonghe, MNRAS, **298**, 677 (1998)
32. P.T. de Zeeuw, MNRAS, **216**, 273 (1985)
33. P.T. de Zeeuw, M. Franx, Annu. Rev. Astron. Astrphys., **29**, 239 (1991)
34. P.T. de Zeeuw, N.W. Evans, M. Schwarzschild, MNRAS, **280**, 903 (1996)
35. S. Durand, H. Dejonghe, A. Acker, A&A, **310**, 97 (1996)
36. E. Emsellem, H. Dejonghe, R. Bacon, MNRAS, **303**, 495 (1999)
37. N.W. Evans, MNRAS, **260**, 191 (1993)
38. N.W. Evans, MNRAS, **267**, 333 (1994)
39. N.W. Evans, R.M. Häfner, P.T. de Zeeuw, MNRAS, **286**, 315 (1997)
40. Ferrari F., Pastoriza M.G., Macchetto F.D., Bonatto C., Panagia N., Sparks W.B., A&A, **389**, 355 (2002)
41. K. Gebhardt, D. Richstone, J. Kormendy, T.R. Lauer, E.A. Ajhar, R. Bender, A. Dressler, S.M. Faber, C. Grillmair, J. Magorrian, S. Tremaine, AJ, **119**, 1157 (2000)
42. K. Gebhardt, T.R. Lauer,J. Kormendy, J. Pinkney, G.A. Bower, R. Green, T. Gull, J.B. Hutchings, M.E. Kaiser, C.H. Nelson, D. Richstone, D. Weistrop, AJ, **122**, 2469 (2001)
43. O.E. Gerhard, MNRAS, **265**, 213 (1993)
44. O.E. Gerhard, P. Saha, MNRAS, **251**, 449 (1991)
45. O.E. Gerhard, G. Jeske, R.P. Saglia, R. Bender, MNRAS, **295**, 197 (1998)
46. J. Gerssen, R.P. van der Marel, K. Gebhardt, P. Guhathakurta, R.C. Peterson,C. Pryor, AJ, **124**, 3270 (2002)
47. P. Goudfrooij, R. de Jong, A&A, **298**, 784 (1995)
48. T.G. Hawarden, A.J. Longmore, S.B. Tritton, R.A.W. Elson, H.G. Corwin, MNRAS, **196**, 747 (1981)
49. C. Hunter, AJ, **80**, 783 (1975)
50. C. Hunter, E. Qian, MNRAS, **262**, 401 (1993)
51. C. Hunter, P.T. de Zeeuw, ApJ, **389**, 79 (1992)
52. Z. Jiang, D. Moss, MNRAS, **331**, 117 (2002)
53. A. Kronawitter, R.P. Saglia, O. Gerhard, R. Bender, A&AS, **144**, 53 (2000)
54. K. Kuijken, ApJ, **446**, 194 (1995)
55. J. Magorrian, D. Ballantyne, MNRAS, **322**, 702 (2001)
56. J. Magorrian, S. Tremaine, D. Richstone, R. Bender, G. Bower, A. Dressler, S.M. Faber, K. Gebhardt, R. Green, C. Grillmair, J. Kormendy, T. Lauer, AJ, **115**, 2285 (1997)
57. A. Mathieu, H. Dejonghe, MNRAS, **303**, 455 (1999)
58. A. Mathieu, H. Dejonghe, X. Hui, A&A, **309**, 30 (1996)
59. M. Matthias, O. Gerhard, MNRAS, **310**, 879 (1999)
60. D. Merritt, T. Fridman, MNRAS, **460**, 136 (1996)
61. D. Merritt, M. Valluri, ApJ, **471**, 82 (1996)

62. A. Pizzella, P. Amico, F. Bertola, L.M. buson, I.J. Danziger, H. Dejonghe, E.M. Sadler, R.P. Saglia, P.T. de Zeeuw, W.W. Zeilinger, A&A, 323, 349 (1997)
63. E.E. Qian, P.T. de Zeeuw, R.P. van der Marel, C. Hunter, MNRAS, **274**, 602 (1995)
64. H.-W. Rix, P.T. de Zeeuw, N. Cretton, R. van der Marel, C. Carollo, ApJ, **488**, 702 (1997)
65. F.H.A. Robijn, P.T. de Zeeuw, MNRAS, **279**, 673 (1996)
66. R.P. Saglia, A. Kronawitter, O. Gerhard, R. Bender, AJ, **119**, 153 (2000)
67. M. Schwarzschild, ApJ, **232**, 236 (1979)
68. M.N. Sevenster, H. Dejonghe, H. Habing, A&A, **299**, 689 (1995)
69. M.N. Sevenster, H. Dejonghe, K. Van Caelenberg, H. Habing, A&A, **355**, 537 (2000)
70. T.S. Statler, AJ, **109**, 1371 (1995)
71. T.S. Statler, H. Dejonghe, T. Smecker-Hane, AJ, **117**, 126 (1999)
72. A. Tomita, K. Aoki, M. Watanabe, T. Takata, S. Ichikawa, AJ, **120**, 123 (2000)
73. H.D. Tran, Z. Tsvetanov, H.C. Ford, J. Davies, W. Jaffe, F.C. van den Bosch, A. Rest, AJ, **121**, 2928 (2001)
74. I. Trujillo, A. Asensio Ramos, J.A. Rubino-Martin, A.W. Graham, J.A.L. Aguerri, J. Cepa, C.M. Gutierrez, MNRAS, 333, 510 (2002)
75. R.P. van der Marel, M. Franx, ApJ, **407**, 525 (1993)
76. R.P. van der Marel, N. Cretton, P.T. de Zeeuw, H.-W. Rix, ApJ, **493**, 613 (1998)
77. P.G. van Dokkum, M. Franx, AJ, **110**, 2027 (1995)
78. P. Vauterin, H. Dejonghe, A&A, **313**, 465 (1996)
79. P. Vauterin, H. Dejonghe, MNRAS, **286**, 812 (1997)
80. E.K. Verolme, P.T. de Zeeuw, MNRAS, **331**, 959 (2002)
81. E.K. Verolme, M. Cappellari, Y. Copin, R.P. van der Marel, R. Bacon, M. Bureau, R.L. Davies, B.M. Miller, P.T. de Zeeuw, MNRAS, **335**, 517 (2002)
82. F. Wernli, E. Emsellem , Y. Copin, A&A, **396**, 73 (2002)

# Dark Matter in Spiral Galaxies

Albert Bosma

Observatoire de Marseille, 2 Place Le Verrier, 13248 Marseille Cedex, France

**Abstract.** In this talk I will discuss several issues concerning the dark matter problem in spiral galaxies. I will give first my version of the state of the debate about cuspy halos in low surface brightness galaxies, and then discuss the situation in bright spirals. I conclude that the dark matter profiles of low surface brightness late-type galaxies do not show much evidence for cuspy halos with NFW type profiles, and that the inner parts of high surface brightness spirals are unlikely to be dominated by dark matter.

## 1 Introduction

Recent developments in cosmology show a rather coherent picture on large scales, in favor of a ΛCDM model of structure formation, with $\Omega_\Lambda$ about 0.7, and $\Omega_{\mathrm{matter}}$ about 0.3. Since $\Omega_{\mathrm{baryon}}$ is about 0.04, there is room for dark matter in the form of a WIMP, but, with $\Omega_{\mathrm{stars}}$ about 0.005 at z = 0, there are a lot of baryons unaccounted for at the present epoch (see e.g. [45] and references therein).

Nevertheless, there is a dark matter "crisis" on small (galaxy) scales, since numerical simulations of cosmological structure formation seem to predict:

- cuspy halos, in apparent conflict with data on low surface brightness (hereafter LSB) galaxies and the inner parts of the Milky Way
- a lot of substructure, in apparent conflict with the number and size of the Milky Way's companions
- an offset in the Tully-Fisher relation, referred to as the angular momentum problem.

Since the majority of people working in the field think that the ΛCDM model of structure formation is the best picture describing the large scale structure, great care has to be exercised to ascertain the validity of the conclusions on the galaxy scale, before concluding about the validity of the ΛCDM model. The reactions to this situation are of several kinds: proponents of the ΛCDM model question the validity of the observations, others examine the role of additional physical processes which could remove material in a dark halo cusp while not questioning the validity of ΛCDM, and yet others examine other theories, such as warm dark matter, self-interacting dark matter, or even MOND.

In this paper I will examine the two issues concerning dark matter on galactic scales on which a lot of work has been done recently, i.e. the dark matter profiles of LSB galaxies, and the question of the dominance of dark matter in the central parts of bright spirals.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 271–278, 2003.

## 2 No Cuspy Halos in LSB Galaxies

### 2.1 Expectations vs. Observations

Dwarf and low surface brightness galaxies are thought to be dark matter dominated, and thus provide a crucial test for the current cosmological numerical simulations, which are predicting the density profiles of dark matter halos. Inner power law slopes for dark halos produced in cosmological simulations of CDM and $\Lambda$CDM models come out to be -1.5 ([33], [17]) or -1.0 (Navarro et al. 1996 [35], hereafter NFW). The latter argue that their NFW profile is universal, and thus can be scaled down to the dwarf galaxy scale (even though these scales are not yet fully modelled directly up till now). For warm dark matter models, similar slopes are found (e.g. [22]).

If one neglects the minor contribution of the stellar and gaseous components in these galaxies, an upper limit can be found for the slope of the density profile of the dark halo directly from inverting the observed rotation curves into a density distribution. It is crucial to get rotation data at the highest spatial resolution possible, but some claims in the literature have been made based not on real data, but on HI rotation curves corrected for beam smearing. Since a clear measurement is better than several arguments, we have collected a large data set using long slit H$\alpha$-spectra, and find that we can exclude the cuspy halos predicted by the current cosmological numerical simulations ([13], [14], [32], [11]). In the last paper, we concentrated on nearby dwarf galaxies, so as to have the highest linear resolution possible. Our results clearly show that the high resolution data favour models with a core, and exclude the steeper slopes required by the NFW and other CDM models.

### 2.2 Reply to Some Criticism

Primack ([39], [40]) severely criticises our work as published in De Blok et al. [13]. Like e.g. Van den Bosch & Swaters [46], he orients the discussion towards asking whether the data are still consistent with NFW profiles, rather than towards trying to find which power law slope fits the data best. In any case, he prunes the data, and eliminates galaxies which have data at poor resolution, edge-on galaxies, etc., so as to retain only a restricted sample. While I disagree with his statistics in the end, I will follow here Primack's arguments, select from our data the galaxies which are well resolved according to his criteria, and see whether his conclusions are justified.

a) *resolution*. Though in [13] and [11] we do discuss the effects of resolution, let me consider only galaxies with **two** independent points inside 1 kpc, the radius at which the discrimination with the NFW model prediction becomes significant. This strict criterion allows me to retain 16 galaxies from the original sample discussed in [13].

b) *edge-on galaxies*. Contrary to Primack's assertions, there is nothing mysterious about their kinematics. In [8] we show that small edge-on galaxies like NGC 100 are transparent, a conclusion confirmed e.g. by [31]. Moreover, our

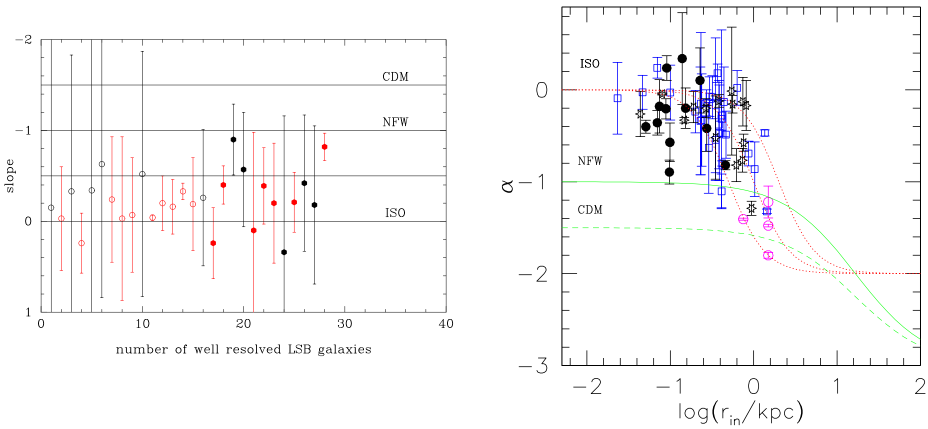



**Fig. 1. Left:** Plots of the slopes for the 16 best resolved galaxies of the De Blok et al. [13] sample (open circles) and the additional 12 galaxies in the De Blok & Bosma [11] sample (filled circles), with 3 $\sigma$ error bars. 17 galaxies are 3 $\sigma$ away from the NFW prediction. **Right:** Value of the inner slope $\alpha$ of the mass-density profiles plotted against the radius of the innermost point on the rotation curve in kpc, for a number of dwarf and low surface brightness galaxies observed by De Blok et al. Open circles and squares are from the sample in [13], stars and filled circles are from the sample in [11]. Overplotted are the theoretical slopes of a pseudo-isothermal halo model (dotted lines) with core radii of 0.5 (left-most), 1 (centre) and 2 (right-most) kpc. The full line represents a NFW model, the dashed line a CDM $r^{-1.5}$ model. Both of the latter models have parameters $c = 8$ and $V_{200} = 100$ km s$^{-1}$, which were chosen to approximately fit the data points in the lower part of the diagram.

spectra do not show low radial velocity wings expected for a non solid body rotation in the central parts, which by itself rules out NFW profiles for these edge-on galaxies. So I retain the 5 edge-on galaxies in the final sample, but note that my conclusions do not depend on their inclusion.

c) *more data*. I add new data from the February 2001 run of [11]. Primack [39] did not dispose of these data, but in [40] chose to ignore them. This brings the total number of galaxies in the sample considered here to 28 galaxies.

With these new selection criteria, I rule out NFW slopes at the 3$\sigma$-level for 17 galaxies, of which 5 edge-on (see Fig. 1). Primack's statement that of the dozen cases probed in [13], about half appear consistent with the cuspy NFW profile is simply not true (I find only 2 out of 12 galaxies he selected having slopes $\leq$ -0.5), and his final conclusion that this data set may be consistent with an inner density profile $\alpha \sim$ -1 but probably not steeper, is not warranted, and not corroborated by the newer data.

### 2.3 Further Observations and Modelling OF Possible Errors

In [11] we already show that data taken by independent observers agree for F561-1, F563-1, F568-3, and UGC 5750. Our results have been corroborated by an independent study by Marchesini et al. [28, 29], who did a further comparison of raw data for 4 other galaxies. Fabry-Perot data from [18] for a few galaxies in common are also in good agreement.

During a February 2002 observing run, I checked again several galaxies previously observed by others, and find good agreement between the raw data. Moreover, I experimented for UGC 4325 with deliberate slit offsets of ± 5", and find very good agreement between the position-velocity curves thus obtained. Such agreement is expected for velocity fields dominated by solid-body rotation. These observations are reported in [12].

Furthermore, to counter objections that systematic errors play a significant role in our results, and always bias the result away from cusps, and in favour of cores, we have simulated the observational analysis using Monte Carlo simulations (cf. [12]. Part of the scatter in the slope $\alpha$ could come from the fact that the "no disk" solution is not strictly true, but this gives a lower limit on $\alpha$. Part of the scatter could come from lopsidedness, part could come from streaming motions, and part from slit offsets, and/or mismatches between the optical and the dynamical centers.

In all these cases, it seems very difficult to erase a peak in the distribution around slope -1 expected if the NFW profile is really valid in LSB galaxies. Only extreme values of the possible biasing effects have to be postulated, i.e. slit offsets of 5" or more in all cases, etc. On the other hand, the good resolution data favour a slope of -0.2, with a scatter of 0.2 which can naturally be explained by a modest amount of every effect considered.

# 3 Spirals

As already remarked by Kalnajs [19], there is not much need for a dark halo in order to explain optical rotation curves of bright spiral galaxies. This has been corroborated for many more optical rotation curves by Kent [20, 21], Athanassoula et al. [2], Buchhorn [9] as reported in Freeman [15], Moriondo et al. [34], and Palunas & Williams [37]. It is thus the extended HI rotation curves which provide the strongest evidence for the need of a dark halo. However, there is no simple way to determine how important the disk is in the central parts, and whether it dominates the potential there. Even so, models based on cosmological simulations (e.g. [36]) predict that the inner parts of spirals are also dark matter dominated. Hence the importance to study spiral galaxies in such a way so as to discriminate between maximum disk or sub-maximum disks.

## 3.1 Importance of Good Data in the Inner Parts

As for the LSB galaxies discussed before, it is very important to get good rotation curve data in the inner parts, so that the decomposition into a mass model with disk (and bulge sometimes), gas and dark halo components can be done correctly. This means in practice that the HI data for the outer parts need to be supplemented with data in the inner parts, if the spatial resolution of the data at 21-cm is inadequate. In the latter case, beam smearing will lower systematically the derived rotation curve. This problem, while thoroughly understood in principle, is still playing an unnecessary large role in the discussion. Partly this is due to the unfortunate circumstance that for late type spirals the differences between the H$\alpha$ and HI rotation curves are not very large, and sometimes so small within the admittedly large errors of both datasets, that some people are misled into assuming that they can be ignored. This is sometimes justified *a posteriori* by taking a high resolution H$\alpha$ spectrum along the major axis, which agrees with the HI rotation curve in some cases, but not in others. Systematic programs are now underway to remedy this situation, using several techniques: long slit spectroscopy of emission lines such as H$\alpha$ (e.g. [13]), 3D imaging using Fabry-Pérot techniques (e.g. [10, 5]), or CO data from e.g. the BIMA interferometer [50].

### 3.2 Importance of Disk Self-gravity

For brighter spirals, the problem of decomposition into mass components and the degeneracy of this procedure becomes crucial. I have discussed this problem several times elsewhere, e.g. [6]. There are several arguments pleading for the disk mass being dominant in the central parts of spiral galaxies, i.e. the constraints posed by spiral structure theory, the constraints imposed by fitting systematic peculiar velocities due to motions induced by spiral arms, and the amplitude of the radial velocity jumps across shocks in barred spirals.

**Swing Amplifier Criteria:** Athanassoula et al. [2] have discussed the application of spiral structure constraints to composite mass models. For each model they examine if swing amplification, based on the mechanism discussed by Toomre [43], is possible. A more graphical description is given in [7]. The physics of the swing amplifier depend on the shape of the rotation curve and on a characteristic X parameter, which in turn depends on the epicyclic frequency $\kappa$, the number of arms m, and the active surface mass density of the disk.

By requiring that the swing amplification of the m = 2 perturbations is possible, one can thus limit the range of mass-to-light ratios to a factor of 2: the lower limit is set by requiring that the disk is massive enough so that amplification of the m = 2 perturbations is just allowed, and the upper limit is set by requiring that amplification of the m = 1 perturbations is just prohibited. Usually the latter condition is fulfilled if one requires a model with maximum disk and a halo with non-hollow core. See [2] for more details.

**Non Circular Motions in Spiral Galaxies:** Already in the M81 HI data obtained with the Westerbork telescope ([41]) the effects of peculiar motions due to the spiral arms are clearly seen. These were modeled with a response calculation by Visser [47], who did not include a dark halo in his models. Lately, Alfaro et al. [1] show a single long-slit spectrum of the galaxy NGC 5427, where the presence of “wiggles” in the position-velocity curve are quite clearly associated with the spiral arms. This is not new, of course, since wiggles were already seen in long slit data of quite a number of galaxies (e.g. [42, 30]), and taken to imply maximum disk (e.g. [15]). It is clear that the presence of such wiggles indicates that the disk is self-gravitating enough to produce them.

Kranz et al. [23] present long-slit data for NGC 4254, a spiral galaxy in the Virgo cluster for which also HI data are available [38]. They try to reproduce the observed velocity perturbations with a stationary gas flow model using the K-band image of the galaxy as input to the evaluation of the disk part of the galactic potential. They find that a maximum disk model produces too large velocity perturbations, and put an upper limit on the disk mass fraction (the mass ratio between a given disk model and the maximum disk model) of 0.8. However, this galaxy is lopsided in the HI, the spiral may be evolving, the small bar in the center of the galaxy might have a different pattern speed than the main spiral pattern, and the inclination may be higher than the authors take it.

In his thesis, Kranz ([24, also 25], reports on a similar analysis for four more cases, and finds a trend that the brightest spirals (those with the highest rotational velocities), seem to have maximum disks, but that towards lower luminosity spirals the relative influence of the dark matter in the inner parts increases. A comparison with the data from Athanassoula et al. [2] shows in fact good agreement with this trend.

**Non Circular Motions in Barred Spirals:** Weiner et al. [48, 49] model in detail the gas flow in the barred spiral NGC 4123, and find that the best fit to the velocity data requires a maximum disk model for the mass distribution. Note that here again the modelling is done as a response calculation for a stationary flow in the potential derived from an optical image. As in the case of NGC 4254, no time evolution has been considered.

Lindblad, Lindblad & Athanassoula [26] have analyzed similar data for the bright barred spiral NGC 1365, and find a relatively good fit with a maximum disk model. Since the rotation curve of NGC 1365 is declining, no dark halo has been included in these models.

### 3.3 Lensing Data

Gravitational lensing is a promising tool for the derivation of mass distributions, since there are no assumptions to make about the messy stellar population contents of the object which does the lensing. Maller et al. [27] discuss the possibility to determine in the case of strong lensing the flattening of the gravitational potential from eventual misalignments of the lensed images and the object doing the lensing. Trott & Webster [44] discuss the case of the lens 2237+0305, for which they combine the lens model with data on the outer rotation curve from VLA HI observations. From their models there is clear evidence for little mass due to the dark halo in the central parts, which are dominated by a bulge-bar system (also [51]). Their interpretation that the disk is not maximal is partly influenced by their inclusion of the bar into the bulge, even though bars are thought to originate in the disk.

For our own Galaxy data on the microlensing towards the bulge-bar system has been used to estimate whether there could be a NFW-like dark matter cusp in the central parts ([3, 4]). Since the lensing data has to be explained by the stars in the bar-bulge system, one can calculate how much the dark matter contributes to the inner rotation curve of the Galaxy, while keeping the constraints set by the situation in the solar neighbourhood. The models for the stellar distribution of the lensing sources slightly underpredict the lensing rate, so that there is hardly any room for dark matter in the central parts of our Galaxy.

## 4 Conclusion

In conclusion, our results show that the predicted steep slopes in the density profiles of LSB galaxies are not observed. The evidence for brighter spirals is that also there dark matter may not be dominant in the central parts of galaxies. This means that the current description by cosmological numerical models is incomplete at small scales, if not incorrect.

### Acknowledgements

Thanks are due to Lia Athanassoula for frequent discussions, and Erwin de Blok and Stacy McGaugh for a fruitful collaboration on LSB galaxies.

## References

1. E.J. Alfaro, E. Perez, R.M. Gonzalez Delgado, M.A. Martos, J. Franco: Astrophys. J. **550**, 253 (2001)
2. E. Athanassoula, A. Bosma, S. Papaiannou: Astron. Astrophys. **179**, 23 (1987)
3. N. Bessantz, O. Gerhard: Mon. Not. R. Astron. Soc. **330**, 591 (2002)
4. J.J. Binney, N.W. Evans: Mon. Not. R. Astron. Soc. **327**, L27 (2001)
5. S. Blais-Ouellette, P. Amram, C. Carignan: Astron. J. **121**, 1952 (2001)
6. A. Bosma: astro-ph/0112080 (2001)
7. A. Bosma: in ASP Conf. Ser. Vol. **182**, Galaxy dynamics, eds. D.R. Merritt, M. Valluri & J.A. Sellwood, (San Francisco: ASP), 339 (1999)
8. A. Bosma, Y.I. Byun, K.C. Freeman, E. Athanassoula: Astrophys. J. **400**, L21 (1992)
9. M. Buchhorn: Ph.D. thesis, Australian National University (1992)
10. R. Corradi, J. Boulesteix, A. Bosma, M. Capaccioli, P. Amram, M. Marcelin: Astron. Astrophys. **244**, 27 (1991)
11. W.J.G. de Blok, A. Bosma: Astron. Astrophys. **385**, 816 (2002)
12. W.J.G. de Blok, A. Bosma, S.S. McGaugh: Mon. Not. R. Astron. Soc.(in press), astro-ph/0212102 (2002)
13. W.J.G. de Blok, S.S. McGaugh, V.C. Rubin: Astron. J. **122**, 2396 (2001)
14. W.J.G. de Blok, S.S. McGaugh, A. Bosma, V.C. Rubin: Astrophys. J. **552**, L23 (2001)
15. K.C. Freeman: in Physics of Nearby Galaxies, Nature or Nurture ?, eds. T.X. Thuan, C. Balkowski & J. Tran Thanh Van (Gif-sur-Yvette: Editions Frontières), 201 (1992)
16. K.C. Freeman: in IAU. Coll. 171, The low surface brightness Universe, eds. J.I. Davies, C. Impey & S. Phillips, (San Francisco: ASP), 3 (1998)

17. T. Fukushige, J. Makino: Astrophys. J. **557**, 533 (2001)
18. O. Garrido, M. Marcelin, P. Amram, J. Boulesteix: Astron. Astrophys. **387**, 821 (2002)
19. A. Kalnajs: in IAU Symp. 100, Internal Kinematics and Dynamics of Galaxies, ed. E. Athanassoula, (Dordrecht: Reidel), 87 (1983)
20. S.M. Kent: Astron. J. **91** 1301 (1986)
21. S.M. Kent: Astron. J. **93** 816 (1987)
22. A. Knebe, J.E.G. Devriendt, A. Mahmood, J. Silk: Mon. Not. R. Astron. Soc. **329**, 813 (2002)
23. T. Kranz, A. Slyz, H.-W. Rix: Astrophys. J. **562**, 164 (2001)
24. T. Kranz: Ph.D. thesis, Heidelberg (2002)
25. T. Kranz, A. Slyz, H.-W. Rix: astro-ph/0212290 (2002)
26. P.A.B. Lindblad, P.O. Lindblad, E. Athanassoula: Astron. Astrophys. **313**, 65 (1996)
27. A.H. Maller, L. Simard, P. Guhathakurta, J. Hjorth, A.O. Jaunsen, R.A. Flores, J.R. Primack: Astrophys. J. **533**, 194 (2000)
28. D. Marchesini, E. D'Onghia, G. Chincarini, C. Firmani, P. Conconi, E. Molinari, A. Zaachei: astro-ph/0107424 (2001)
29. D. Marchesini, E. D'Onghia, G. Chincarini, C. Firmani, P. Conconi, E. Molinari, A. Zaachei: astro-ph/0202075 (2002)
30. D.S. Mathewson, V.L. Ford, M. Buchhorn: Astrophys. J. Suppl. **81**, 413 (1992)
31. L.D. Matthews, K. Wood: Astrophys. J. **548**, 150 (2001)
32. S.S. McGaugh, V.C. Rubin, W.J.G. de Blok: 2001, Astron. J. **122**, 2381 (2001)
33. B. Moore, T. Quinn, F. Governato, J. Stadel, G. Lake: Mon. Not. R. Astron. Soc. **310**, 1147 (1999)
34. G. Moriondo, R. Giovanelli, M.P. Haynes: Astron. Astrophys. **346**, 415 (1999)
35. J.F. Navarro, C.S. Frenk, S.D.M. White: Astrophys. J. **462**, 563 (1996)
36. J.F. Navarro: astro-ph/9807084 (1998)
37. P. Palunas, T.B. Williams: Astron. J. **120**, 2884 (2000)
38. B. Phookun, S.N. Vogel, L.G. Mundy: Astrophys. J. **418**, 113 (1993)
39. J.R. Primack: astro-ph/0112255 (2001)
40. J.R. Primack: astro-ph/0205391 (2002)
41. A.H. Rots, W.W. Shane: Astron. Astrophys. **31**, 245 (1974)
42. V.C. Rubin, D. Burstein, W.K. Ford Jr, N. Thonnard: Astrophys. J. **289**, 81 (1985)
43. A. Toomre: 'What amplifies the spirals'. In *The Structure and Evolution of Normal Galaxies*, eds. S.M. Fall & D. Lynden-Bell (Cambridge: Cambridge Univ. Press), pp. 111-136, (1981)
44. C.M. Trott, R.L. Webster: Mon. Not. R. Astron. Soc. **334**, 621 (2002)
45. M.S. Turner: astro-ph/0207297 (2002)
46. F. van den Bosch, R.A. Swaters: Mon. Not. R. Astron. Soc. **325**, 1017 (2001)
47. H.C.D. Visser: Astron. Astrophys. **88**, 159 (1980)
48. B. Weiner, J.A. Sellwood, J.H. van Gorkom, T.B. Williams: Astrophys. J. **546**, 916 (2001)
49. B. Weiner, J.A. Sellwood, T.B. Williams: Astrophys. J. **546**, 931 (2001)
50. T. Wong: Ph.D. Thesis, University of Berkeley (2000)
51. H.K.C. Yee: Astron. J. **95**, 1331 (1988)

# A SAURON View of Galaxies

Ellen K. Verolme[1], Michele Cappellari[1], Glenn van de Ven[1], P. Tim de Zeeuw[1], Roland Bacon[2], Martin Bureau[3], Yanick Copin[4], Roger L. Davies[5], Eric Emsellem[2], Harald Kuntschner[6], Richard McDermid[1], Bryan W. Miller[7], and Reynier F. Peletier[8]

[1] Sterrewacht Leiden, Leiden, The Netherlands
[2] Centre de Recherche Astronomique de Lyon, Saint–Genis–Laval, Lyon, France
[3] Department of Astronomy, Columbia University, New York, USA
[4] Institut de Physique Nucléaire de Lyon, Villeurbanne, France
[5] Physics Department, University of Oxford, Oxford, UK
[6] European Southern Observatory, Garching, Germany
[7] Gemini Observatory, La Serena, Chile
[8] Department of Physics and Astronomy, University of Nottingham, Nottingham, UK

**Abstract.** We have measured the two-dimensional kinematics and line-strength distributions of 72 representative nearby early-type galaxies, out to approximately one effective radius, with our panoramic integral-field spectrograph SAURON. The resulting maps reveal a rich variety in kinematical structures and linestrength distributions, indicating that early-type galaxies are more complex systems than often assumed. We are building detailed dynamical models for these galaxies, to derive their intrinsic shape and dynamical structure, and to determine the mass of the supermassive central black hole. Here we focus on two examples, the compact elliptical M32 and the E3 galaxy NGC4365. These objects represent two extreme cases: M32 has very regular kinematics which can be represented accurately by an axisymmetric model in which all stars rotate around the short axis, while NGC4365 is a triaxial galaxy with a prominent kinematically decoupled core, with an inner core that rotates about an axis that is nearly perpendicular to the rotation axis of the main body of the galaxy. Our dynamical models for these objects demonstrate that two-dimensional observations are essential for deriving the intrinsic orbital structure and dark matter content of galaxies.

## 1 The SAURON Project

The formation and evolution of galaxies is one of the most fundamental research topics in astrophysics. A key question in this field is whether early-type galaxies form very early in the history of the universe or are gradually built up by mergers and the infall of smaller objects. The answer to this problem is closely tied to the distribution of intrinsic shapes, the internal dynamics and linestrength distributions, and the demography of supermassive central black holes.

In the few past decades, it has become clear that ellipticals, lenticulars, and spiral bulges display a variety of velocity fields and linestrength distributions. Two-dimensional spectroscopy of stars and gas is essential when attempting to derive information on the intrinsic structure. For this reason, we have built a panoramic integral-field spectrograph, SAURON ([1]), which provides large-scale two-dimensional kinematic and linestrength maps in a single observation.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 279–285, 2003.

We commissioned SAURON on the 4.2m William Herschel Telescope on La Palma in 1999. In low-resolution mode, the spectrograph combines a large field-of-view ($33'' \times 41''$) with a pixel size of $0\farcs94$. When the seeing conditions are good, the high-resolution mode, with a pixel size of $0\farcs28$, allows zooming in on galactic nuclei. SAURON observes in the spectral range of 4810–5340 Å, which contains the gaseous emission lines H$\beta$ and [OIII] and [NI], as well as a number of stellar absorption features (Mg$b$, Fe, H$\beta$). The instrumental dispersion is $\sim$100 km/s. Between 1999 and 2002, we have used SAURON to observe a carefully-selected representative sample of 72 ellipticals, lenticulars and Sa bulges, distributed over a range of magnitudes, ellipticities, morphologies and environments ([29]).

We have finalized the data reduction, have accurately separated the emission- and absorption lines, have calibrated the line-strength measurements, and have in hand maps of the stellar and gaseous kinematics and linestrengths for all 48 E and S0 objects, with those for the spirals to follow soon. The maps reveal many examples of minor axis rotation, decoupled cores, central stellar disks, non-axisymmetric and counter-rotating gaseous disks, and unusual line-strength distributions ([2, 29]). We have also developed new methods to spatially bin the data cubes to a given signal-to-noise ([4]), and to quantify the maps with Fourier methods ([5, 17]). This allows accurate measurements of, e.g., the opening angle of the isovelocity contours and of the angle between the direction of the zero-velocity contour and the minor axis of the surface brightness distribution ([5]), enabling various statistical investigations of the entire sample of objects.

## 2 Dynamical Models

We are constructing detailed dynamical models which fit all kinematics and eventually even observations of the stellar line-strengths of the galaxies in the SAURON survey. We do this by means of Schwarzschild's ([18]) orbit superposition method, which was originally developed to reproduce theoretical density distributions (e.g., [12, 14, 18–20]), and was subsequently adapted to incorporate observed kinematic data in spherical and axisymmetric geometry ([6, 9, 11, 16]). We have implemented a number of further extensions including the ability to deal with a Multi-Gaussian Expansion of the surface brightness distribution ([3, 8, 13]). We have also shown that the large data sets that are provided by instruments such as SAURON can be modelled without any problems ([26]).

Recently, we completed the non-trivial extension to the software that allows inclusion of kinematic measurements in triaxial geometry ([27]). As in the axisymmetric case, observational effects such as pixel binning and point-spread-function convolution are taken into account. The chaotic orbits are dealt with in the 'standard' way (see [24]), and the line-of-sight velocity profile is used to constrain the models. In the next two sections, we describe two applications in more detail, one in axisymmetry and the other for a triaxial intrinsic shape.

## 3 Axisymmetric Models for M32

We applied our axisymmetric modeling software to the nearby compact E3 galaxy M32 ([26]). By complementing the SAURON maps (Fig. 1) with high-resolution major axis stellar kinematics taken with STIS ([10]), the models are constrained at both small and large radii, which allows us to measure an accurate central black hole mass $M_{\mathrm{BH}}$, stellar mass-to-light ratio $M/L$, *and* inclination $i$. The left panels of Fig. 2 show the dependence of $\Delta\chi^2$, which is a measure of the discrepancy between model and data, on $M_{\mathrm{BH}}$, $M/L$ (in solar units, for the $I$-band) and $i$. The inner three contours show the formal 1, 2 and $3\sigma$-confidence levels for a distribution with three degrees of freedom. The black hole mass and mass-to-light ratio are constrained tightly at $M_\bullet = 2.5 \times 10^6 M_\odot$ and $M/L = 1.8 M_\odot/L_\odot$, and the inclination is constrained to a value near $70° \pm 5°$. The right panels of the same figure show similar contours, but now for a data-set consisting of the STIS-kinematics together with four slits extracted from the SAURON-data. In this case the constraints on all three parameters, but most notably on the inclination, are much less stringent. This demonstrates that two-dimensional observations are essential to gain insight into the intrinsic structure of galaxies.

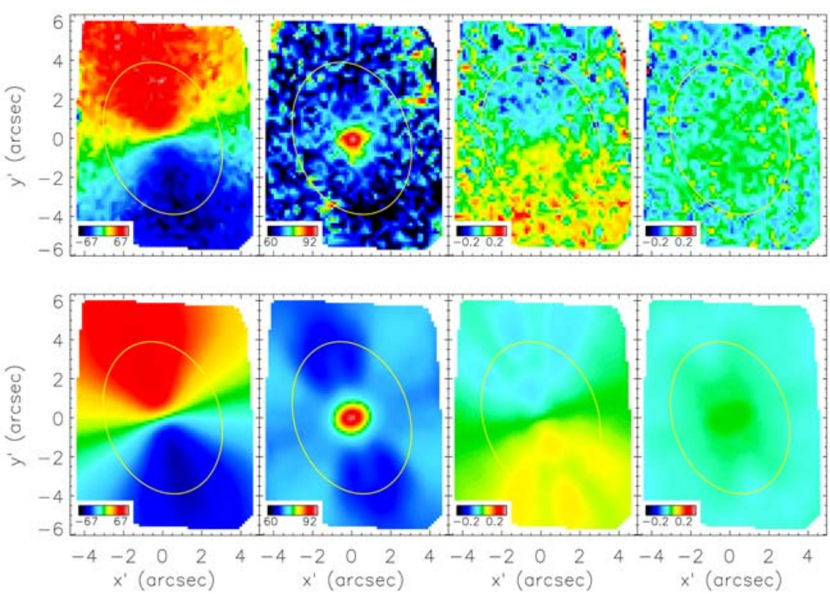



**Fig. 1.** Top panels: the SAURON kinematic maps for M32. From left-to-right: the mean velocity, velocity dispersion and Gauss–Hermite parameters $h_3$ and $h_4$, which measure the first and second order deviations of the line-of-sight velocity distribution from a Gaussian shape. Bottom panels: idem, but now for the best-fit axisymmetric dynamical model with $I$-band $M/L = 1.8 M_\odot/L_\odot$, $M_{\mathrm{BH}} = 2.5 \times 10^6 M_\odot$, and $i = 70°$.

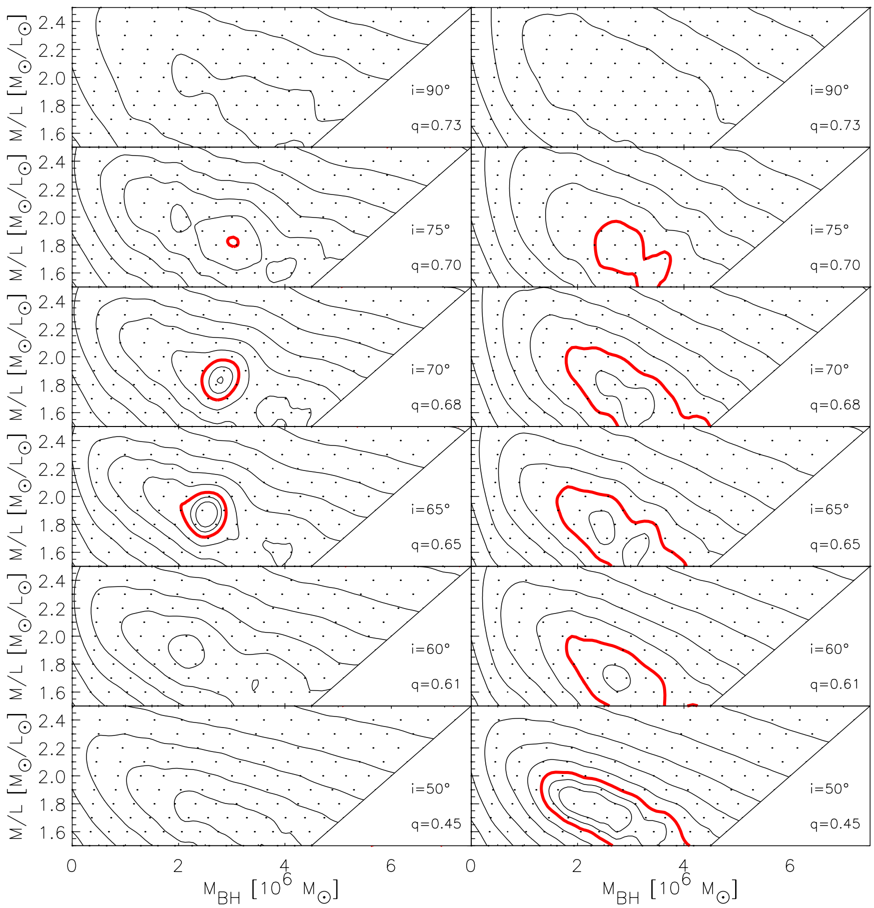



**Fig. 2.** Dynamical models for M32 ([26]). The panels show contours of the goodness-of-fit parameter $\Delta\chi^2$ as a function of the central black hole mass $M_{\rm BH}$, the stellar mass-to-light ratio $M/L$ and the inclination $i$. Each dot represents a specific axisymmetric dynamical model. The intrinsic flattening $q$ of the models is indicated in the lower-right corner of each panel. The models are constrained by `STIS` kinematics along the major axis ([10]) together with two-dimensional observations obtained with `SAURON` in its high resolution mode ([29]). The inner three contours represent the formal $1, 2$ and $3\sigma$-confidence levels for a distribution with three degrees of freedom. *Left panels*: model fits to a data set consisting of the `STIS`-data and the full `SAURON` field. Tight constraints are placed on the central black hole mass and mass-to-light ratio, as well as on the allowed range of inclinations. *Right panels*: the $\Delta\chi^2$ for models that were constrained by four extracted slits from the $9'' \times 11''$ `SAURON` field (major and minor axis, and at $\pm45°$, as in [11]) and the `STIS` data. This shows that the traditional kinematic coverage provides almost no constraint on $i$, and that the resulting uncertainties on the inferred values of $M/L$ and $M_{\rm BH}$ are correspondingly larger.

## 4 The Triaxial Galaxy NGC 4365

The upper panels of Fig. 3 show the stellar kinematics in the central $30'' \times 60''$ of the giant elliptical galaxy NGC4365, derived from two SAURON pointings ([7]). The velocity field clearly shows a prominent decoupled core in the inner $3'' \times 7''$

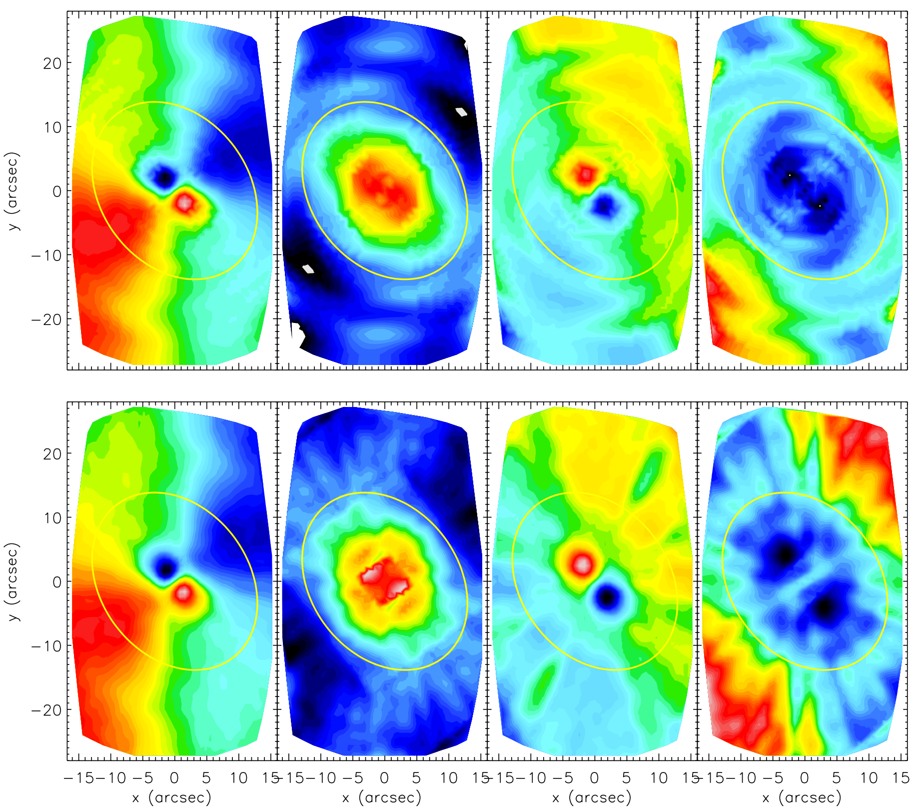



**Fig. 3.** Observations and dynamical models for the E3 galaxy NGC4365. *Top panels*: from left to right, the stellar velocity field, velocity dispersion, and Gauss–Hermite moments $h_3$ and $h_4$, as observed with SAURON. The maps are based on two semi-overlapping pointings, sampled at $0.''8 \times 0.''8$, and were constructed via a kinemetric expansion to provide the best representation of the data that is consistent with an intrinsically triaxial geometry (e.g., point-antisymmetry for the $V$ and $h_3$ maps, cf. [5]). The original maps can be found in ([7]). The amplitude of the velocity field is about 60 km/s, the peak velocity dispersion is 275 km/s, and the contours in the $h_3$ and $h_4$ maps range between $\pm 0.10$. The decoupled core measures $3'' \times 7''$. *Bottom panels*: idem, but now for a dynamical model with average intrinsic axis ratios $p = 0.93$ and $q = 0.69$ (triaxiality parameter $T = (1-p^2)/(1-q^2) \sim 0.22$), observed from a direction defined by the viewing angles $\vartheta = 85°$ and $\varphi = 15°$. This model reproduces all the main characteristics of the observations ([28]).

(cf. [23]). It has a rotation axis which lies $82° \pm 2°$ away from that of the body of the galaxy, which rotates around its long axis. Such a structure is possible when the shape is intrinsically triaxial because of the presence of orbits that have net mean streaming around either the long or the short axis.

The globular cluster system of NGC 4365 shows evidence for an intermediate age population ([15]). The SAURON linestrength maps, however, indicate a predominantly old stellar population ([7]), suggesting that the observed kinematic structure may have been in place for over 12 Gyr and the galaxy is in stable triaxial equilibrium. We therefore applied our developed modeling software to this case, to investigate whether it is possible to reproduce all the kinematic data in detail, and to constrain $M/L$ and the intrinsic shape and orbital structure.

We represented the observed surface brightness distribution of NGC 4365 by a Multi-Gaussian Expansion which accurately fits the observed radial variation of ellipticity, the boxyness of the isophotes, and the modest isophotal twisting. We derived the deprojected density by assuming that each of the constituent Gaussian components is stratified on similar concentric triaxial ellipsoids. The three Euler angles that specify the orientation of the ellipsoids can be chosen freely. For each choice, we computed a library of 4000 orbits, obtained from 20 energy shells with 200 orbits each, covering the four major orbit families, and including orbits from minor families and chaotic orbits. As the spatial resolution of the SAURON measurements is modest, we did not consider the effect of a central black hole. The preliminary results indicate that the quality-of-fit parameter $\Delta\chi^2$ varies quite significantly with $M/L$ and the parameters defining the intrinsic shape. The lower panels of Fig. 3 show the predictions of one model that fits the data well. This illustrates that the software works, and shows that NGC 4365 is indeed consistent with a triaxial equilibrium shape.

In principle, best-fit values of the shape parameters, the direction of observation, and the mass-to-light ratio can be determined by a systematic investigation of the parameter space, just as was done for M32. For triaxial systems this is a very time-consuming effort, but a first-order guess of the galaxy parameters can be obtained by using other, simpler, schemes (see, e.g., [21, 22, 25]). Our detailed dynamical modeling software can then be used to explore this more restricted parameter range. Work along these lines is in progress, and will make it possible to deduce, e.g., the intrinsic properties of the kinematically decoupled cores seen in many of these systems. Inclusion of higher spatial resolution data will allow accurate measurement of the mass of the central black hole.

## 5 Concluding Remarks

We have presented two examples of recent results from our program to construct detailed axisymmetric and triaxial dynamical models for galaxies in the SAURON representative survey of nearby ellipticals, lenticulars and Sa bulges. The panoramic SAURON observations tighten the constraints on the possible orientation of a galaxy considerably. The extension of the modeling software to triaxial shapes including kinematic constraints works, and that it will help us gain significant insight into the structure of early-type galaxies.

## References


1. Bacon R., et al., 2001, MNRAS, 326, 23
2. Bureau M., et al., 2002, in ASP Conf. Ser., 273, 53
3. Cappellari M., et al., 2002, ApJ, 578, 787
4. Cappellari M. & Copin Y., 2003, MNRAS, in press
5. Copin Y. et al., 2001, EDPS Conf. Ser. in Astron. & Astrophys., eds F. Combes, D. Barret, F. Thévenin, 289 (astro-ph/0109085)
6. Cretton N., de Zeeuw P.T., van der Marel R.P., Rix H.–W., 1999, ApJS, 124, 383
7. Davies R.L., et al., 2001, ApJ, 548, L33
8. Emsellem E., Monnet G., Bacon R., 1994, A&A, 285, 739
9. Gebhardt K., et al., 2003, ApJ, in press (astro-ph/0209483)
10. Joseph C.L., et al., 2001, ApJ, 550, 668
11. van der Marel R.P., Cretton N., de Zeeuw P.T. & Rix H.-W., 1998, ApJ, 493, 613
12. Merritt D., Fridman T., 1996, ApJ, 460, 136
13. Monnet G., Bacon R., Emsellem E., 1992, A&A, 253, 366
14. Poon M.Y., Merritt D., 2002, ApJ, 568, 89
15. Puzia T., Zepf S.E., Kissler–Patig M., Hilker M., Minniti D., Goudfrooij P., 2002, A&A, 391, 453
16. Rix H., de Zeeuw P.T., Cretton N., van der Marel R.P., Carollo C.M., 1997, ApJ, 488, 702
17. Schoenmakers R.H.M., Franx M., de Zeeuw P.T., 1997, MNRAS, 292, 349
18. Schwarzschild M., 1979,, ApJ, 232, 236
19. Schwarzschild M., 1993, ApJ, 409, 563
20. Siopis C., Kandrup H. E., 2000, MNRAS, 319, 43
21. Statler T.S., 1994a, ApJ, 425, 458
22. Statler T.S., 1994b, ApJ, 425, 481
23. Surma P. & Bender R., 1995, A&A, 298, 405
24. Terzic B., Hunter C., de Zeeuw P.T., 2001, in *Stellar Dynamics: from Classic to Modern*, eds L.P. Osipkov & I.I. Nikiforov, (St. Petersburg State University, Russia), 303
25. van de Ven G., Hunter C., Verolme E.K., de Zeeuw P.T., 2003, MNRAS, submitted
26. Verolme et al., 2002, MNRAS, 335, 517
27. Verolme E.K., Cappellari M., van de Ven, G., de Zeeuw P.T., 2003, MNRAS, submitted
28. Verolme E.K., et al. 2003, in prep.
29. de Zeeuw et al., 2002, MNRAS, 329, 513

# Photometric Properties of Karachentsev's Mixed Pairs of Galaxies

Alfredo Franco-Balderas, Deborah Dultzin-Hacyan,
and Héctor M. Hernández-Toledo

Instituto de Astronomía, UNAM. Apartado Postal 70-264, C.P. 04510 México D.F., México.

**Abstract.** We present multicolor broad band BVRI photometry for a sample of 42 mixed morphology binary galaxies taken from the "Karachentsev Catalogue of Isolated Pairs of Galaxies" (KPG). Images were obtained with 0.84m and 1.5m telescopes of the Observatorio Astronómico Nacional, San Pedro Mártir, Baja California, México, operated by the Instituto de Astronomía, UNAM. Our goal is to identify and isolate some structural and photometric properties of disk galaxies and elliptical galaxies at different stages of interaction.

## 1 Review

The mid to late 70's saw an astronomical debate that led to the recognition that gravitational interaction is an important factor in galactic evolution affecting directly properties such as size, morphological type, luminosity, star formation rate, and mass distribution (Sulentic, [20]; Larson & Tinsley,[14] ; Stocke, [19]). According to current popular models of galaxy formation, galaxies are assembled through a hierarchical process of mass aggregation, dominated either by mergers (Kauffmann, White & Guiderdoni, [11]; Baugh, Cole & Frenk, [3]), or by gas accretion (Avila-Reese, Firmani & Hernández [2]; Avila-Reese & Firmani, [1]). In the light of these models, the influence of environment factors and interaction phenomena in the shaping and star formation of the disks is natural, at least for a fraction of the present-day galaxy population.

Pairs of galaxies occupy an initial position in the spectrum of galaxy populations and are used to measure, in approximate way, the mass of the components as well as to know the form of the gravitational potential on the basis of the morphology of the components and its evolution in time. For binary galaxies, current ideas suggest that most physical pairs are morphologically concordant, that is, with components showing similar initial star formation and angular momentum properties. However, the number of (E+S) pairs ($\sim$128) in The Catalogue of Isolated Pairs of Galaxies in the Northern Hemisphere (KPG, Karachentsev, [8]) means that, for a flux limited sample ($m < 15.7$), more than two out of every ten pairs are of the (E+S) type, suggesting that a considerable number of them must be physical binaries. The KPG was done under a criterion of strict isolation that excluded, as far as possible the optical pairs. Redshift information, available for the whole (E+S) sample, suggests that most of them are likely to be physically proximate. Digital Sky Survey images show that most of them

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 286–293, 2003.

have visible signs of disturbance; bridges, tails, common envelopes and distortions that are regarded as evidence for gravitational interaction. In addition, statistical studies indicate that a high fraction ($\sim 65$ %) show an enhancement in the optical and FIR emission (Xu & Sulentic, [21]; Hernández-Toledo et al., [6]). This enhancement is interpreted as a by-product of interaction-induced star formation activity in physical binaries (Rampazzo & Sulentic, [16]). This is at odds with the current models of galaxy formation.

We present multicolor broad band BVRI photometry for a sample of 40 elliptical-spiral (E+S) binary galaxies taken from the "Karachentsev Catalogue of Isolated Pairs of Galaxies" (KPG). Images were obtained with 0.84m and 1.5m telescopes of the Observatorio Astronómico Nacional, San Pedro Mártir, Baja California, México, operated by the Instituto de Astronomía, UNAM. Images were calibrated using standard stars from Landolt's [13] list.

Our goal is to identify and isolate some structural and photometric properties of disk galaxies and elliptical galaxies at different stages of interaction.

## 2 Interactions, Mergers and Evolution

Most E+S pairs show signs of disturbance like bridges, tails, geometric and morphologic distortion, and in some cases nuclear activity, which are evidence of gravitational interactions.

Karachentsev [8] identified three basic interaction classes (AT, LI and DI) that describe the pairs which show obvious signs of interaction. AT class identifies pairs with components in a common luminous halo with a symmetric, amorphous or shredded, asymmetric (sh) structure (Fig. 1a). LI pairs show evidence of tidal bridges (br), tails (ta) or both (br+ta) (Fig. 1b). DI pairs show evidence of structural distortion in one (1) or both (2) components (Fig. 2a). We add to this sequence NI for pairs with no obvious morphological distortion (Fig. 2b).

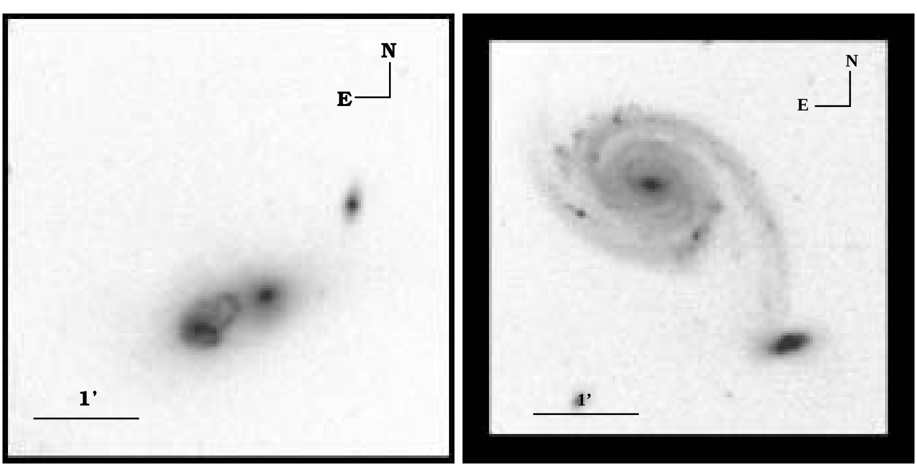



**Fig. 1.** Interaction classes. a)Left: KPG83 (AT pair). b)Right: KPG591 (LI pair)

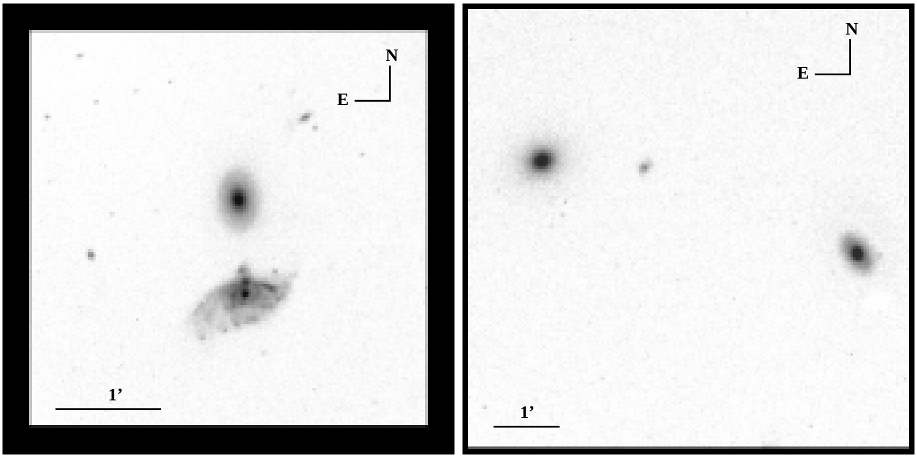



**Fig. 2.** Interaction classes. a)Left: KPG29 (DI pair). b)Right: KPG38 (NI pair)

The order AT-LI-DI-NI can be regarded as a sequence from strongest to weakest evidence for tidal distortion or, alternatively, most to least dynamically evolved (interpreting a common envelope as a sign of extensive dynamical evolution in pairs).

While mergers between comparable mass galaxies may be responsible for some of the most dramatic extragalactic events, minor mergers of small mass satellites may play a subtle but nonetheless critical role in the evolution of galaxies. Satellites are commonly found in the vicinity of normal galaxies (Zaritsky et al., [22]).

## 3 Surface Photometry

Mixed pairs of galaxies are excellent laboratories for the study of interaction in galaxies because they represent a set of objects in which we see the effects of the interaction of a rich gas object (S member) on the presence of a relatively clean disturber (E member).

The surface brightness profile of a galaxy is produced by the spatial distribution of the stars as well as the spatial distribution of the dust. To discuss the optical morphology (that could be modified by the presence of bars, spiral arms, rings, etc) and its relationship to the global photometric properties, the final results for each pair are presented in the form of a mosaic (see Fig. 3). Each mosaic includes: 1)A $B$-band image, 2) A $B$-band filtered image, 3) Surface brightness and color profiles, and 4) Its correspondent geometric profiles (radial $\epsilon$, P.A. and $A_4$ coefficient). In most of the cases, foreground stars in the field have been removed.

We are proposing to use our filtered images in combination with the estimated geometric and surface brightness profiles to look for morphological features in more detail. The morphologic evaluation can be done in three parts: 1.- Visual

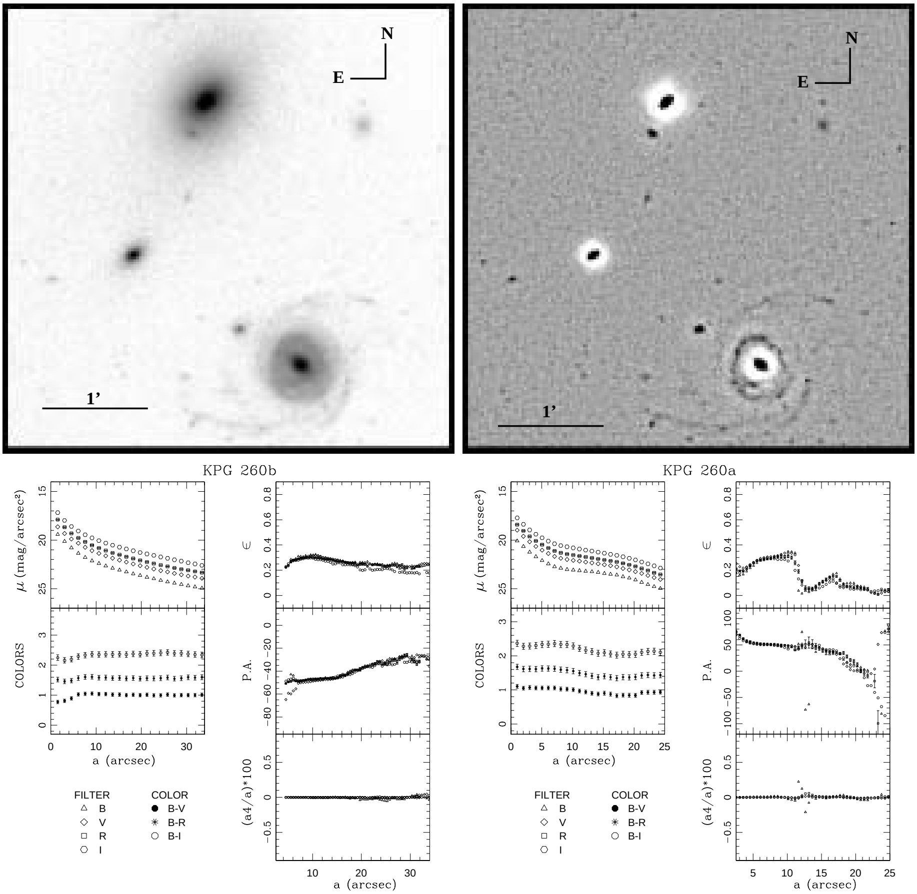



**Fig. 3.** KPG260 Mosaic. Top left: B-band image. Top right: B-band filtered image. Bottom left: Surface Brightness Profile and Geometric Parameters for KPG260A (west). Bottom right: Surface Brightness Profile and Geometric Parameters for KPG260B (east).

identification of false pairs E+S. 2.- Evaluation of the geometric parameters for each component. 3.- Evaluation of all characteristics, mainly those related to structures like spiral arms, shells, rings, bars and large regions of stellar formation, presumably associated with the interactions.

HST has revealed that the cores of some ellipticals (as well as the central regions of some spirals) have an excess of light relative to the de Vaucouleurs law fit. The excess light and rapid decline are evidence for central black holes. Some bright ellipticals also deviate from the de Vaucouleurs law in their outer regions, with a surface brightness in excess of the expected fit. There is evidence that mergers, capture of material and tidal disturbances from other galaxies play an important role in the final structure of elliptical galaxies.

On the other hand, spiral galaxies have bulges that are very similar to elliptical galaxies. In the disk, the brightness decreases approximately exponentially, with a characteristic scale of length for each galaxy. Barred galaxies, like non-barred galaxies, have exponential disks. There are two types of bars: Flat bars have almost constant surface brightness along the bar, that is, they have a much shallower decline than the disk. Exponential bars, in contrast, have the same scale of length as the disk.

Measurements of surface brightness profiles are essential for quantitative investigations of galaxy morphology, decomposition of bulge and disc, studies of galaxy structure and stellar populations, and measurements of dust distribution. Important parameters like the size, the distribution of mass, the star formation rate, as well as the magnitude of nuclear activity can be determined with the help of our observations.

## 4 Deviations from Perfect Ellipticity

The intensity of an elliptical galaxy can be expressed as:

$$I = I_0 + A_n * \sin(n * \phi) + B_n * \cos(n * \phi) \tag{1}$$

where $\phi$ measures the position angle of the major axis. The $A_n$ and $B_n$ coefficients for n > 1 represent the amplitudes of the deviations form perfect ellipticity, which are typically around 1 %. The quantity :

$$\frac{a_4}{a} = \frac{\sqrt{1-\epsilon} * B_4}{a * |\frac{dI}{da}|} \tag{2}$$

(see Bender & Möllenhoff, [4]), were $a$ is the length of the isophote's semimajor axis, forms a dimensionless measure of the diskiness of the isophote, indicate whether a galaxy is boxy ($a_4/a < 0$) or disky ($a_4/a > 0$). See Fig. 4.

The traditional view on the formation and evolution of giant elliptical galaxies is that they are very old stellar systems and all formed very early at a redshift of more than two (Searle, Sargent & Bagnuolo [18]). Alternatively, hierarchical theories of galaxy formation predict that massive galaxies were assembled relatively late in many generations of mergers of disk-type galaxies or smaller subunits and mass accretion. It has been argued by Kauffman [9] and Kauffman & Charlot [10] that this merger scenario is consistent with observations of elliptical galaxies at different redshifts. Naab & Burkert [15] performed a large set of collisionless N-body simulations (taking into account the stellar and dark matter component) of merging disk-galaxies with mass ratios of 1:1, 2:1, 3:1, and 4:1. They show that 1:1 merger remnants rotate slowly, are supported by anisotropic velocity dispersions, have significant minor-axis rotation, and show predominantly boxy isophotes in good agreement with observations of bright giant ellipticals. 3:1 and 4:1 remnants are isotropic, fast rotators, show a small amount of minor-axis rotation, and have disky isophotes in perfect agreement with observations of faint fast rotating giant elliptical galaxies. 2:1 remnants show intermediate properties.

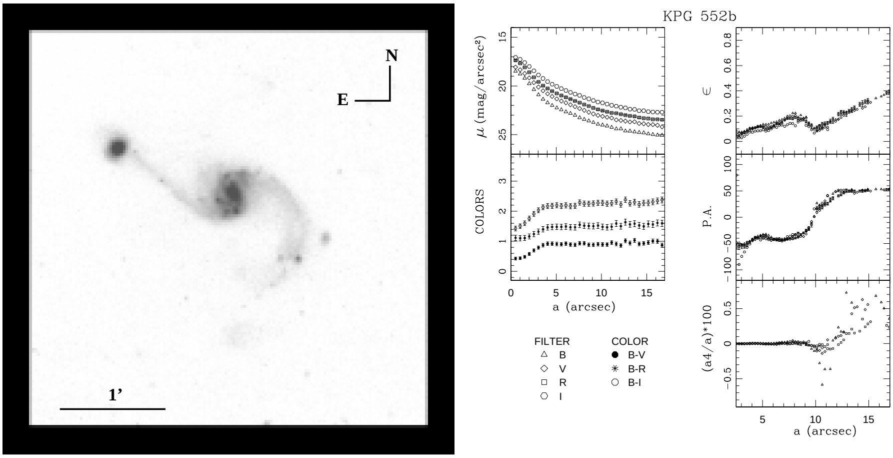



**Fig. 4.** Example of an elliptical galaxy with isophotal twist and external disky structure. Left: B-band image of KPG552. Right: Surface Brightness Profile and Geometric Parameters for KPG552B (west).

Projection effects lead to a large spread in the data in good agreement with our observations.

The frequency and spatial distribution of disky and boxy ellipticals in pairs E+S could provide interesting information on the frequency of equal- and unequal mass mergers in different environments

## 5 Isophotal Twist

The major axis position angle (PA) gives the orientation of the galaxy in the sky: it is measured counterclockwise from north to the major axis. Variation of PA with radius, or twist, may be an indication of triaxiality for elliptical galaxies with no axis of rotational symmetry. Isophotal twist may be common in spirals. Such twist may also be interpreted as an indication of the presence of small bars, rings or nonaxisymmetric bulges in the central regions. Variations in ellipticity are associated with a isophotal twist.

## 6 The Holmberg Effect

Holmberg [7] compared the photographic colors of paired galaxies and found a significant correlation between the colors of pair components. The physical explanation of the Holmberg effect is complex, it has been interpreted as reflecting a tendency for similar types of galaxies to form together (morphological concordance), a possible reflection of the role of local environment in determining galaxy morphology, but alternatively, it can presumably also reflect mutually induced star formation (Kennicutt et al., [12]) in physical pairs.

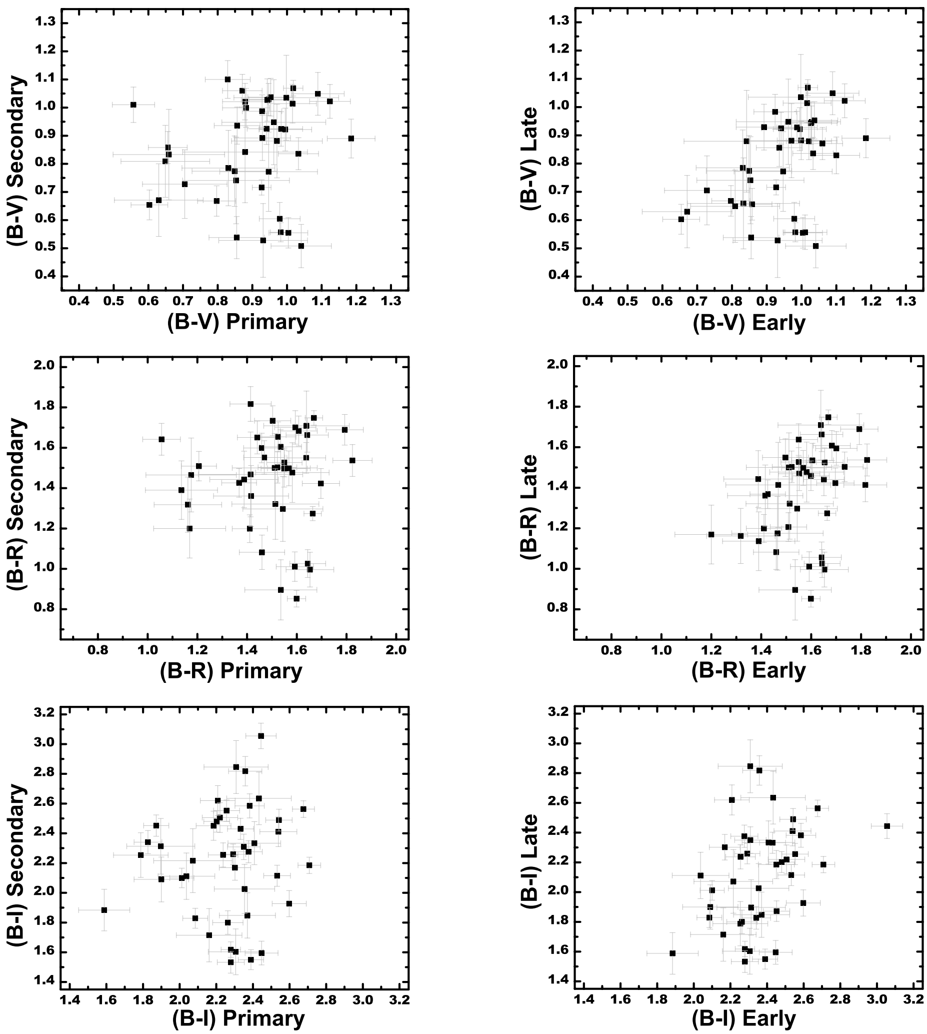



**Fig. 5.** Holmberg Effect. Left column:color index of the fainter member in a E+S pair (secondary component) referred to the color index of the brighter (primary component). Right column: color index of the S member versus the color index of the E member.

Figure 5 show the Holmberg Effect for our sample. The color correlation between pair components in mixed pairs is poor in any plot, contrary to the results of Demin [5]. Any tendency, if present, could be explained by the intrinsic scatter in the $(B-V)^0_T$ - Morphological Type correlation as reported by Roberts & Haynes [17]. However the evidence is not conclusive due to the magnitudes are not corrected by galactic extinction, K-correction and inclination correction.

## 7 Summary

Until this moment, we conclude that:

• An important number of E+S pairs are misclassified ($\sim$ 25 %). The pairs not classified as true (E+S) pairs are primarily of two types: 1) disky pairs composed of spiral and lenticular components and 2) early pairs composed of elliptical and lenticular components. Both of these classes raise interesting questions about galaxy formation because of the discordant star formation and angular momentum properties of the components.

• Most spirals in mixed pairs have redder colors than normal spirals [(B-V) $>$ 0.6] which could be related with formation of dust by an event of starburst.

• On the other hand, the ellipticals in mixed pairs have redder colors than normal ellipticals too [(B-V) $>$ 0.9] which could be related with transference of gas and dust belonging to the spiral galaxy.

• In E+S pairs, "grand design" spirals are common and luminous, whereas a few low-luminosity S members exhibit flocculent spiral structure.

• Boxy ellipticals tend to be more luminous than disky ones, in E+S pairs.

• Possibly boxy ellipticals are formed by mergers of (mostly) ancestral objects. On the other hand, disky ellipticals may have been formed from a single proto-galaxy, or from the merger of mainly gaseous ancestral objects.

• Morphological distortion is common in all the sample.

## References

1. Avila-Reese V. & Firmani C., 2000, Rev. Mex. Astron. Astrofís. 36, 23.
2. Avila-Reese V., Firmani C. & Hernández X., 1998, ApJ 505, 37.
3. Baugh C. M., Cole S. & Frenk C. S., 1996, MNRAS 283, 1361.
4. Bender R. & Möllenhoff C., 1987, A&A 177, 71.
5. Demin V. V., Zasov A. V., Dibai E.A., Tomov A.N., 1984, SvA 28, 367.
6. Hernández-Toledo H. M., Dultzin-Hacyan D., González J. J. & Sulentic J. W., 1999, AJ 118, 108.
7. Holmberg E., 1958, Lund Medd. Astron. Obs. Ser. II, 136, 1.
8. Karachentsev I. D., 1972, Catalogue of Isolated Pairs of Galaxies in the Northern Hemisphere, Comm. Spec. Ap. Obs. 7, 1.
9. Kauffman, L.H., 1996, magr.meet.
10. Kauffman, L.H., & Charlot 1998, pfg..conf.
11. Kauffmann G., White S. D. M. & Guiderdoni B., 1993, MNRAS 264, 201.
12. Kennicutt R. C., Roettiger K. A., Keel W. C., van der Hulst J. M. & Hummel E., 1987, AJ 93, 1011.
13. Landolt A. U., 1992, AJ 104, 340.
14. Larson R. B. & Tinsley B. M., 1978, ApJ 219, 46.
15. Naab T. & Burkert A., 2001, ApJ 555, 91.
16. Rampazzo R. & Sulentic J. W., 1992, A&AS 259,43.
17. Roberts M. S. & Haynes M. P. 1994, ARA&A 32, 115.
18. Searle L., Sargent W. L. W. & Bagnuolo W. G., 1973, ApJ 179, 427.
19. Stocke J. T., 1978, AJ 83, 348.
20. Sulentic J. W., 1976, ApJS 32, 171.
21. Xu C. & Sulentic J. W., 1991, ApJ 374, 407.
22. Zaritsky D., Kennicutt R.C. & Huchra J. P., 1994, ApJ 420, 87.

# Spline Histogram Method for Reconstruction of Probability Density Functions of Clusters of Galaxies

Dmitrijs Docenko[1] and Kārlis Bērziņš[2]

[1] Institute of Astronomy, University of Latvia, Raina blvd 19, Riga LV-1586, Latvia; `dima@latnet.lv`

[2] Ventspils International Radio Astronomy Center, Akademijas laukums 1-1503, Riga LV-1050, Latvia; `kberzins@latnet.lv`

**Abstract.** We describe the spline histogram algorithm which is useful for visualization of the probability density function setting up a statistical hypothesis for a test. The spline histogram is constructed from discrete data measurements using tensioned cubic spline interpolation of the cumulative distribution function which is then differentiated and smoothed using the Savitzky-Golay filter. The optimal width of the filter is determined by minimization of the Integrated Square Error function.

The current distribution of the TCSplin algorithm written in f77 with IDL and Gnuplot visualization scripts is available from www.virac.lv/en/soft.html.

## 1 Introduction

Whenever one makes a physical measurement one obtains a discrete result, starting from spatial measurements and ending with classification of some set of objects by some quantity. Particular measured value follows from the statistical properties of the system strictly following the probability distribution function, hereafter PDF. The PDF, in its turn, is determined by the physical properties of the system. If a measured data set is statistically complete then its PDF contains information about the system's physical properties. The PDF shows a character of unimodal or multimodal systems. It is natural to assume that the PDF of unimodal physical systems contain only one global maximum and several maxima indicate the multimodality of a data set. Therefore the shape of the PDF allows one to classify the measured data points, e.g. to find structure in case of positional measurements.

In statistics it is widely accepted to use histograms as the PDF approximations. It is also well known that ordinary histograms being dependent on two free parameters (bin size and its location) give very subjective results. Many methods have been developed trying to solve this problem [5]. However, most of them are still dependent on some parameters in a non-objective manner.

Generally, the probability density estimation methods can be divided into two main groups: parametric and non-parametric. The first ones assume some definite type of the PDF function (e.g. Gaussian or their superposition) and try to find the best-fit parameters for it. A good such example is the KMM mixture modelling algorithm [2]. A main disadvantage of these methods is that not all

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 294–301, 2003.

data sets can be well fitted with any chosen function. Rather often the real PDF of physical system has significant difference from a chosen best-fit function, and in many cases it is not known *a priori* what function it should be at all.

Non-parametric methods try to construct PDF estimates as compromise of two opposite demands. First, the estimate should be as close as possible to the measured PDF. Second, statistical noise due to a finite volume of the selection should be filtered out. There are several ways how to do it.

It is possible to minimize a functional that is a sum of two terms – the statistical noise and the one increasing with a difference between data points and the PDF estimate (Vondrak's method) [21]. Unfortunately, there is still one free coefficient responsible for the smoothing degree. This coefficient is not determined in any automated way and usually is found from good-looking conditions. Another method is to convolve the initial guess of the PDF with some kernel (kernel methods) for data smoothing. Also in this case there remains a free coefficient – the kernel width, that is responsible for smoothing of the function in an "optimal way", besides the result is weakly dependent on the chosen kernel shape [5], [20].

There is, however, a method that allows one to choose an optimal smoothing width: the PDF should not be over-smoothed and lose its true features, and the noise level should be diminished as far as possible on the other hand. This method is described e.g. in [20]. Its main idea is to define an Integrated Square Error ($ISE$) function that shows the difference between the real PDF and its estimate, and then to minimize it. The $ISE$ function method is implemented for kernel methods in e.g. [15], and results are encouraging. However, the $ISE$ function itself is often rather noisy.

In this paper we propose another approach to estimate PDF of a given one-dimensional data set in automatic and optimal way. This is the spline histogram method. We have found that the tensioned cubic splines are suitable for this task and the corresponding algorithm has been called TCSplin. The current version of the TCSplin code is freely available in the internet, it is also included in the CD-ROM.

For demonstration purposes in this article we will use the spline histogram algorithm to find a "well determined" redshift structure of galaxies within clusters Abell 2256 and Abell 3626.

This paper has the following structure. The spline histogram algorithm will be discussed in section 2. Bootstrapping simulations, discussed in section 3, help to evaluate errors of the PDF estimates. In section 4 the spline histogram application to data sets of clusters of galaxies A2256 and A3526 will be shown as examples. Finally, some concluding remarks will be given in section 5.

## 2 The Spline Histogram Algorithm

The spline histogram method is a non-parametric approach for reconstruction of probability density function underlying statistical selection. It was first discussed in [4] as one of possible methods to detect substructure in clusters of galaxies.

Recently it was further developed in [9] and these results are summarized in this paper.

From spectroscopic observations we obtain redshifts of galaxies in clusters. Let us denote redshift of the $i$th galaxy by $z_i$ and order them ascendentally ($z_i \leq z_{i+1}$). Next step is to construct a step-like cumulative distribution function (CDF) obtained purely from observational data: $F_{obs}(cz) = N(z_j < z)/N_{gal}$, where $N(z_j < z)$ is a number of galaxies with redshift smaller than $z$, and $N_{gal}$ is a total number of detected galaxies, $c$ is the speed of light. At this stage we assume that the data set is statistically complete being representative of the physical situation. The PDF $f(x)$ by definition is the derivative of $F_{obs}(cz)$ in respect to $cz$. If CDF is constructed as shown before then $f(x)$ is a sum of Dirac $\delta$-functions.

In the spline histogram method the points $z_i$ with ordinates $F_{obs}(cz_i)$ are consequently connected by non-decreasing smooth analytical spline $S(cz)$. After construction of $S(cz)$ the latter is analytically differentiated leading to the PDF estimate $\hat{f}(cz)$. This procedure guarantees that the obtained continuous PDF is in agreement with the discrete distribution of the data points. The PDF contains all initially observed information about the cluster and it has a lot of noise as a consequence. To diminish the noise, $\hat{f}(cz)$ has to be smoothed.

Trying to utilize usual cubic splines for interpolation of the CDF, one encounters the problem that they will generally have negative derivative intervals if both first and second derivatives in the data points are put to be equal. Although there is an infinite amount of possibilities how to construct a smooth continuous spline that has non-smooth first derivative at data points.

We have found that tensioned cubic splines (hereafter TCS) nicely fit all spline histogram needs. They are defined such that the cubic polynomial spline length between two data points is minimal, and only the interpolating function and its first derivative are continuous in data points. Also in this case sometimes a derivative of the TCS is negative. Then in order to exclude a non-physical decreasing of the CDF estimate, we use non-tensioned splines increasing accordingly the spline length within these regions.

To reduce the statistical noise, the algorithm has been symmetrized. For the same purpose there was added a possibility to unite close points in the data set, that would otherwise give unphysical high PDF peaks. Nevertheless the resulting PDF is noisy and due to this the next step of the algorithm is a smoothing procedure.

In our case the noise is seen as narrow high peaks in the PDF arising from high CDF derivatives between close data points. The Savitzky-Golay filters [16] have been chosen for the smoothing remembering that PDF construction without any *a priori* knowledge about the system character was one of the main reasons for developing the spline histograms. These filters locally conserve first moments of the smoothed function. The remaining problem is to define the optimal width of the filter such that it reduces the noise but not over-smoothes the real PDF features.

This is done using the Integrated Square Error ($ISE$) function [20]:

$$ISE(\hat{f}(cz)) = \int\limits_{cz_{min}}^{cz_{max}} \left(\hat{f}(cz) - f(cz)\right)^2 d(cz), \tag{1}$$

where $f(cz)$ is a true PDF underlying the observed selection, and $\hat{f}(cz)$ is a PDF estimated from the observed data, i.e. the smoothed spline histogram in our case. As we are using digital filters to smooth the data, this should be rewritten for case of discrete points. It follows from the theory [18], [9], [20] that quantity $P(h)$ will have minimum for the same smoothing width $h$ as $ISE(\hat{f}(x))$:

$$P(h) = \sum_{i=1}^{N} \left(\hat{f}(cz_i)^2 - 2\hat{f}(cz_i) + 2C_0^{(h)}\right), \tag{2}$$

where $C_0^{(h)}$ are the smoothing filter zeroth coefficients, and it was taken into account that $\sum_{i=1}^{N} \hat{f}(cz_i) = N$. In contrary to the equation defining the $ISE$ function, $P(h)$ can be easily calculated from the data.

The filter width that gives the minimal $P(h)$ value is the optimal one because the corresponding deviation between the true and estimated PDFs is also minimized. As a result the spline histogram is obtained but it says nothing about the remaining statistical noise level in it. To find it out we use a bootstrapping technique described in the next section.

## 3 Simulated Data Analysis

Simulated data are produced and analysed as follows. Using the obtained spline histogram as a true PDF, we generate the same amount $N$ of random numbers. Then from this selection we compute another smoothed spline histogram. Repeating this sufficient number of times (say 100), one can calculate the average of the simulated spline histograms and its scattering. It is useful to characterize the scattering by the distribution quartiles. The upper quartile shows that the estimated PDF has 75% probability to be below it. For the lower quartile, accordingly, this probability is 25% (see Fig. 1).

To estimate the quality of the approximation, the first moments of several simulated distributions were computed and compared with the original values (see Table 1). It can be seen that the average values are the same within statistical error bars ($1\sigma$), whereas the standard deviations are about 10% larger than the original values because of the smoothing effect. For Gaussian distributions the asymmetry and excess are significantly different from zero, although in non-Gaussian cases they are rather close to original values.

Dependence of the smoothing size on the selection volume was also analysed. From theoretical considerations [18], [20], [15], we know that the optimal smoothing size depends on the selection volume $N$ in the following way: $h_{opt} \propto N^{-1/5}$. Analysing different volume random number selections for the same initial distribution, we have empirically found that for our algorithm $h_{opt} \propto N^{-0.195}$, that shows an excellent agreement with the theoretical prediction.

**Table 1.** Moments of the initial distribution from simulations of the 500 point selection

| **Gaussian distribution** | Average | St.Dev. | Asymmetry | Excess |
|---|---|---|---|---|
| General distribution | 0.500 | 0.089 | 0.000 | -0.006 |
| Average from simulations | 0.499 | 0.093 | -0.486 | 2.287 |
| St.Dev. from simulations | 0.004 | 0.003 | 0.109 | 0.355 |
| **2 equal dispersion Gaussians** | Average | St.Dev. | Asymmetry | Excess |
| General distribution | 0.475 | 0.190 | 0.190 | -0.718 |
| Average from simulations | 0.474 | 0.192 | 0.153 | -0.676 |
| St.Dev. from simulations | 0.009 | 0.005 | 0.068 | 0.104 |
| **2 different dispersion Gaussians** | Average | St.Dev. | Asymmetry | Excess |
| General distribution | 0.565 | 0.190 | -0.112 | -0.836 |
| Average from simulations | 0.564 | 0.192 | -0.165 | -0.704 |
| St.Dev. from simulations | 0.009 | 0.004 | 0.064 | 0.105 |

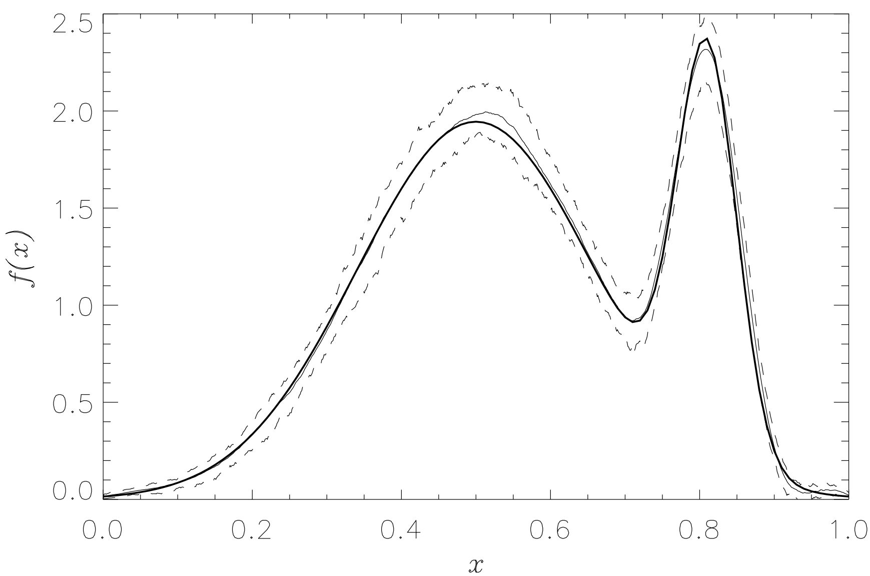



**Fig. 1.** Result of the simulated distribution analysis. Original PDF is shown by the thick solid line, the thin solid line represents the average value of 100 smoothed spline histograms using 500 point selection each, and the lower and upper dashed lines are the first and third quartiles, respectively.

## 4 Galaxy Cluster Data Analysis

As example we show the implementation of the algorithm on two clusters of galaxies – Abell 2256 and 3526.

Abell 2256 is a rich regular cluster at $z \approx 0.06$ ($\alpha \approx 17^h03.7^m$, $\delta \approx +78°43'$, equinox 2000.0 [1]). It has similar properties to the Coma cluster (similar X-ray luminosities, both have optical and X-ray substructure and a radio halo), but is situated approximately 2.5 times farther.

A2256 has been previously studied in x-rays, optical and radio by several authors, e.g. [10], [6], [7]. It is accepted and understood, that Abell 2256, being one of the best studied clusters of galaxies, exhibit complex inner structure.

Result of the implementation of the TCSplin algorithm to the data of [10], consisting of 89 galaxy redshift measurements, is shown in Fig. 2. From the figure one can obviously see that the cluster is unrelaxed and has strongly non-Gaussian velocity PDF. Most likely it consists of two or more merging parts that currently are undergoing a final stage of unification.

Centaurus cluster A3526 ($z \approx 0.011$, $\alpha \approx 12^h 48.9^m$, $\delta \approx -41°18'$, equinox 2000.0 [1]) has been extensively studied, as it is a nearby rich cluster of galaxies. It is intermediate between Coma and Virgo clusters in richness and in distance and has richness class 1 or 2 (e.g. [17]). Centaurus is irregular in appearance, like Virgo. The cluster core has two apparent centres of concentration, one being centred on NGC 4696 and the other being 0.5° further east ([12], [3]).

Extensive study of this cluster is made in [8], [13] and [14]. The research included determination of redshifts for 259 galaxies and photometry for 329 galaxies within 13° field centred on the cluster, and the following analysis of data. The bimodal galaxy velocity distribution and extensive substructure in both subclusters have been found. Mean heliocentric velocities and line-of-sight dispersions of two main cluster components, within 3° of the cluster centre, are 3041 and 586 km sec$^{-1}$ (denoted Cen30), and 4570 and 262 km sec$^{-1}$ (denoted Cen45), respectively. The projected distributions of members of each component overlap on the sky. Other small galaxy groups also have been found in this study.

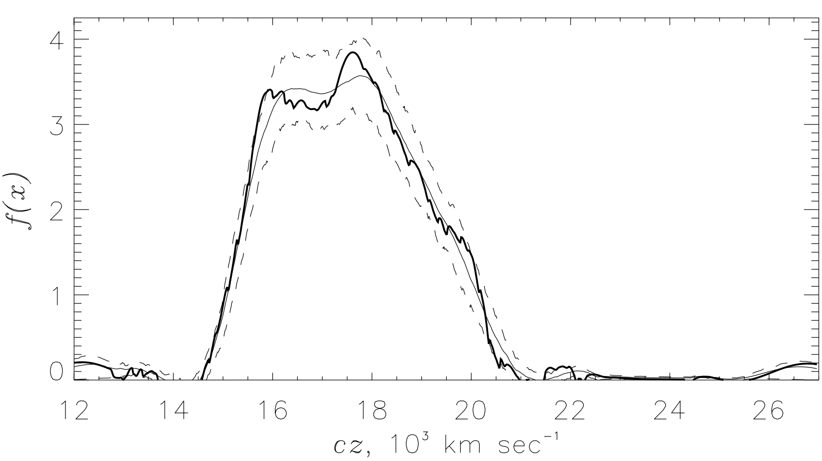



**Fig. 2.** The spline histogram of A2256. The meaning of lines is the same as in Fig. 1

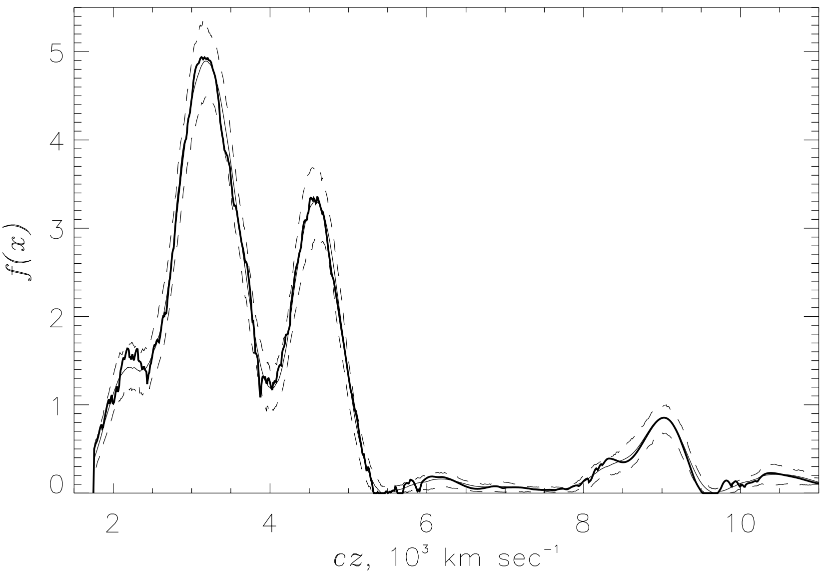



**Fig. 3.** The spline histogram of A3526. The meaning of lines is the same as in Fig. 1

Recently bimodality of the cluster has been confirmed in [19]. The authors used it to test a non-parametric method of the PDF estimation proposed by [11] and the same two main features of the cluster were noticed.

Our result of processing the data set of [8] is shown in Fig. 3. We find the same two main structures as in the original analysis. Clearness of these features demonstrates the quality of the algorithm. Shape of each of these components is close to Gaussian indicating their relatively relaxed state. Besides that the spline histogram shows additional left "shoulder" of the Cen30 group at around $cz \approx 2100$ km sec$^{-1}$. This probably is one of separate groups noticed by [14]. Possibly this as well as those features around 6200 and 8300 km sec$^{-1}$ are not spatially real but just the redshift space caustics artefacts.

We see that a direct implementation of the algorithm leads to a good estimate of the PDF of clusters of galaxies. The only difference is the dispersions of the group velocities that are overestimated due to our PDF smoothing. One should keep that in mind and calculate the dispersion directly from the original data if needed.

## 5 Concluding Remarks

This paper has demonstrated the usefulness of the spline histogram algorithm in statistical studies of 1D data sets. It has all advantages over the well known ordinary histogram approach estimating the probability density functions. In principle the spline histograms may be expanded to higher dimensional cases but that introduces higher effect of the sampling noise. Unfortunately enlargement of a data set size does not necessarily guarantee larger signal to noise ratio. More generally it is dependent on the distribution character.

The latest version of the spline histogram algorithm TCSplin code is freely available online from http://www.virac.lv/en/soft.html. Presently it is written in f77 with IDL and Gnuplot visualization scripts.

### Acknowledgments

DD is grateful to the European Physical Society EWTF that has provided a travel grant and the LOC of Research centre of Astronomy of Academy of Athens for possibility to participate in the Workshop. DD and KB also thank Bernard Jones for the fruitful discussions. The work of KB was supported by grant No.01.0024.4.1 of the Latvian Council of Sciences.

## References

1. G.O. Abell, H.G. Corwin, Jr, and R.P. Olowin: ApJSS, **70**, 1 (1989)
2. K.A. Ashman, C.M. Bird, and S.E. Zepf: AJ, **108**, 2348 (1994)
3. N.A. Bahcall: ApJ, **193**, 529 (1973)
4. K. Berzins: in *Generation of Cosmological Large-Scale Structure* ed. by D.N. Schramm and P. Galeotti (Kluwer Academic Publishers, 1997) pp.283-288
5. K. Berzins: *Substructure of clusters of galaxies: Methodology*, Master Thesis, University of Copenhagen and University of Latvia, Copenhagen, Riga (1998), http://www.virac.lv/papers/kberzins_msc.tar.gz
6. U.G. Briel, et al.: A&A, **246**, L10 (1991)
7. T.E. Clarke, and T.A. Ensslin: astro-ph/0106137 (2001)
8. R.J. Dickens, M.J. Currie and J.R. Lucey, MNRAS **220**, 679 (1986)
9. D. Docenko: *Spline histograms and their application to analysis of clusters of galaxies* (in Latvian), Master Thesis, University of Latvia, Riga (2002), http://www.virac.lv/papers/ddocenko_msc.pdf
10. D.G. Fabricant, S.M. Kent, and M.J. Kurtz: ApJ, **336**, 77 (1989)
11. D. Fadda, E. Slezak, and A. Bijaoui: A&ASS **127**, 335 (1998), astro-ph/9704096
12. A.R. Klemola: AJ, **74**, 804 (1969)
13. J.R. Lucey, M.J. Currie and R.J. Dickens: MNRAS, **221**, 453 (1986)
14. J.R. Lucey, M.J. Currie and R.J. Dickens: MNRAS, **222**, 427 (1986)
15. A. Pisani: MNRAS **265**, 706 (1993)
16. W.H. Press, S.A. Teukolsky, W.T. Vetterling, and B.P. Flannery: *Numerical Recipes in FORTRAN*, Cambridge University Press, Cambridge (1992)
17. A. Sandage: ApJ, **183**, 731 (1973)
18. B.W. Silverman: *Density Estimation for Statistics and Data Analysis*, Chapman & Hall, London (1986)
19. P. Stein, H. Jerjen, and M. Federspiel: A&A **327**, 952 (1997), astro-ph/9707211
20. R. Vio, G. Fasano, M. Lazzarin, and O. Lessi: A&A, **289**, 640 (1994)
21. J. Vondrak: Bulletin of Astronomical Institute of Czechoslovakia, **20**, 349 (1969)

# Stars Close to the Massive Black Hole at the Center of the Milky Way

Nelly Mouawad[1], Andreas Eckart[1], Susanne Pfalzner[1], Christian Straubmeier[1], Rainer Spurzem[2], Reinhard Genzel[3], Thomas Ott[3], and Rainer Schödel[3]

[1] I. Physikalisches Institut, University of Cologne, Zülpicher Str.77, 50939 Cologne, Germany
[2] Astronomisches Rechen-Institut, Mönchhofstr. 12 bis 14, 69120 Heidelberg, Germany
[3] Max-Planck-Institut für extraterrestrische Physik, Giessenbachstr., 85748 Garching, Germany

**Abstract.** Recent measurements of stellar velocities ([5], [7]) and variable X-ray emission [3] near the center of the Milky Way have already provided the strongest case for the presence of a super-massive black hole in our Galaxy. Information on the enclosed mass and stellar number density counts, in the central stellar cluster of the Galaxy, now allows to derive realistic potentials to study stellar orbits. We present the results of calculations using a $4^{th}$-order Hermite integrator. They provide valuable additional information on the three dimensional distribution and dynamics of the He-Stars. We also discuss the importance of Newtonian peri-astron shifts for stellar orbits in the central cluster and how future observations with infrared interferometers (LBT, VLTI, Keck) [6] will help to improve our understanding of the dynamics and distribution of the stars in this region.

## 1 Introduction

Stellar proper motions, radial velocities and accelerations obtained with high angular resolution techniques over the past decade have convincingly proven the presence of 3 million solar masses in the center of our Galaxy. This mass is associated with the compact radio source Sagittarius A* and currently represents the best candidate for a super massive Black Hole. In this gravitational potential at the center of the Milky Way, the stars show large orbital velocities. In the central arcsecond those can be observed as proper motions via repeated imaging at the highest possible resolution.The largest velocities in the vicinity of SgrA* are several 1000 km/s with a maximum $\geq$5000 km/s for the early type star S2. The location at the maximum stellar velocity dispersion and the center of gravitational force, as determined from the orbital accelerations of the two stars S1 and S2 agree to within less than 100 mas with the position of the compact radio source SgrA*. These accelerations were consistent with bound orbits but still allowed for a wide range of possible orbits ([9], [7]).

For the first time, and after 10 years of observations with the MPE speckle camera SHARP at the ESO NTT and the new adaptive optics CONICA/NAOS at the UT4 of the Very Large Telescope (VLT), we are now able to determine an

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 302–309, 2003.

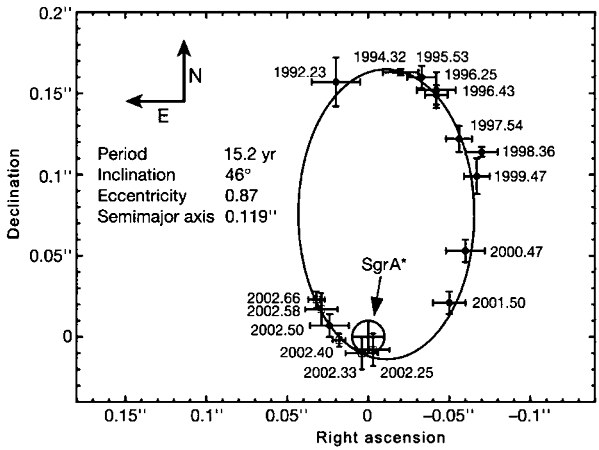



**Fig. 1.** Orbit of S2, relative to the position of the compact radio source SgrA* (large cross and circle, denoting the ±10 mas uncertainties of the infrared-radio astrometry). The data obtained with the MPE SHARP camera at the NTT and the VLT UT4 NACO adaptive optics are shown. The projection of the best-fitting Kepler orbit is shown as a continuous curve with its main parameters listed adjacent to the orbit.

orbit for the star S2, currently closest to the compact radio source SgrA* (semi-major axis = 4.62 mpc), these data trace an almost complete orbit (Fig. 1) and are well fitted by a bound, highly elliptical Keplerian orbit ($\epsilon$=0.87), with an orbital period of 15.2 years, requiring an enclosed point mass of $(3.7\pm1.5)\times10^6$ $M_\odot$ [15]. This star passed through peri-center in April 2002, at a distance of only 17 light hours from the radio source, when it moved at $\geq$5000 km/s.

## 2 Mass Estimate for the Inner Cusp

### 2.1 Stellar Density

With the highest spatial resolution observations presently available in the near-infrared (50 - 60 mas), spatial scales from light hours to a few light years can be probed. The new CONICA/NAOS data were estimated from direct and crowding corrected $K_s\leq18$ stellar counts of in annuli centered on the position of SgrA*. They clearly confirm the presence of a local stellar concentration, a cusp centered on SgrA*, indicated earlier by the SHARP/NTT and KECK data ([4], [1]). To estimate the cusp mass, we were able to fit the combined SHARP and CONICA stellar count data with a superposition of several Plummer models of the form: $\rho(r)=\rho(0)[1+(\mathrm{r/R})^2]^{-\alpha/2}$ ($\alpha$=5), with different densities $\rho(0)$, and different core radii R. Fig. 2 shows three different fits, where the dotted curve shows a fit for the inner cusp with a PLummer model of a core radii R=0.55" and a spatial density $\rho(r)$=4.35*$10^7$ $M_\odot$ $\mathrm{pc}^{-3}$. The solid line shows the sum of that initial model with a further inner Plummer model of R=0.135" and $\rho(r)$=6.5*$10^8$ $M_\odot$ $\mathrm{pc}^{-3}$. The average fit is shown in the large dotted curve. It is similar to the solid

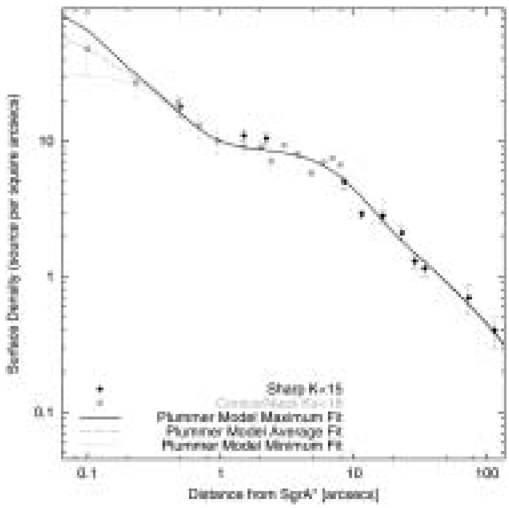



**Fig. 2.** A Plummer Model fit to the surface density of stars as a function of distance. The grey, filled circles represent the CONICA/NAOS data for $K_s \leq 18$. The darker, filled diamonds represent the SHARP/NTT data for stars with $K_s \leq 15$, and scaled upward by a factor of 5 in order to match the fainter CONICA/NAOS counts [8]. The small dotted curve and the solid curve represent the minimum and maximum fit to the data with our model, respectively. The big dotted middle curve represents the average fit.

curve but with a smaller density of $\rho(r)$=3.25*$10^8$ $M_\odot$ $pc^{-3}$ of the additional R=0.135" component.

## 2.2 The Enclosed Mass

From these parameters, we were able to derive the enclosed mass as a function of the separation from the Galactic Center. Under the assumption that the mass to light ration of the inner cluster is comparable to that of the outer cluster (M/L∼ 2μm), the mass present at a distance of 0.55" from the BH was evaluated to be between 5000 $M_\odot$ and 6100 M⊙. Figure 3 shows 3 different curves in solid, dash-dotted and grey dotted. Similar to Fig. 2 they correspond to maximum, mean, and minimum fit, respectively. Using a fourth order Hermite integrator derived from the one used in high-accuracy $N$-body simulations ([2], [13] for the first introduction of the Hermite scheme see [10]), and adapting the mid-value of our models, we computed ,for an S2 like orbit [15] , the trajectory of a star through the extended mass and around the BH . The Hermite scheme allows a fourth order accurate integration based on only two time steps. For that it requires the analytic computation of the time derivative of the gravitational force; therefore the use of Plummer model superpositions as they are used here is very convenient. Further studies with more general density distributions should be undertaken , because they may influence the precise value of the periastron shift. Also it should be noted that this integration is purely classical so any relativistic periastron shifts are not taken into account.

The resulting retrograde periastron shift amounts to a value of ~1.7 arcmins per orbital revolution which is few times smaller than the relativistic prograde periastron shift [12]. Figure 4 below gives us different periastron shift values for different cusp masses, i.e. mass to light ratios.

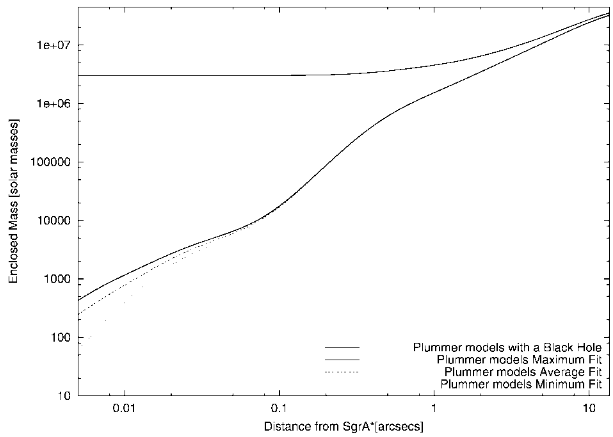



**Fig. 3.** Mass distribution in the central 10 pc of the Galaxy. In black a model fit to the data [8] resulting from the Plummer models fit to the stellar number density data and a $3\times10^6$ $M_\odot$ black hole. The 3 lower curves show a the stellar enclosed mass only. They clearly show the contribution of the inner cusp to the overall mass. The grey dotted curve represents a minimum of 5000 $M_\odot$ present inside a sphere of 0.55". The dash-dotted curve is the same as for the first curve with an additional mass of 550 $M_\odot$ present in 0.135". The solid curve is similar to the dash-dotted curve but with a larger additional mass of 1100 $M_\odot$

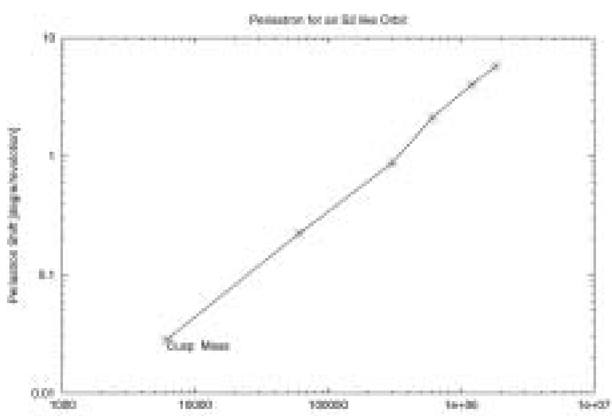


**Fig. 4.** Variation of the periastron shift with the inner cusp mass.

## 3 Estimate of the Line of Sight Positions of 13 Helium Stars

An enigmatic case in the central cluster of our Galaxy is the presence of the He I emission line stars and other early type stars (e.g in IRS16/13 complexes). They are confined in a projected radius on the sky not larger than about 400 mpc. Genzel et al. [8] show that most of them are on tangential orbits and seem to have a projected clockwise rotation on the sky. A determination of the spatial positions of these stars, will help to clarify the way they were formed, how they heat the gas and dust in the central parsec, and - potentially - how they influence the accretion stream onto the central arcsecond. We studied 13 (of the 29 known) He I emission-line stars, for which Genzel et al. [8] had given 3D velocities, and 2D projected separations from SgrA* (see also the list in [11]). With these parameters we compute the trajectories of the stars using a fourth order Hermite integrator, and assuming a potential derived from the the plummer model fit described above (Fig. 3).

Assuming that these stars are orbiting inside a sphere of 400 mpc or inside a sphere of 200 mpc, we can then determine the maximum line of sight ranges for the different stars. Figure 5 shows the results obtained for the 400 mpc case (dotted) and the 200 mpc case (solid black). For some of the stars the line of

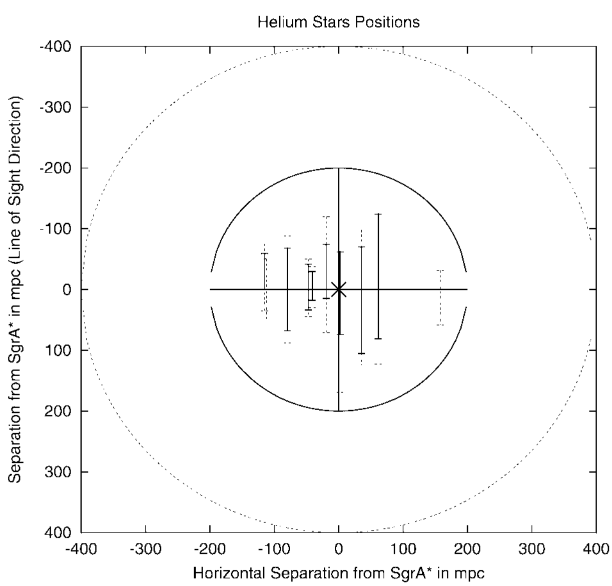



**Fig. 5.** The dash-dotted exterior circle is the projection of a sphere of 400 mpc of radius, Black solid interior circle is the projection of a sphere of 200 mpc. Dotted red bars are possible ranges along the line of sight positions of the He-stars with orbits that stay well within a sphere of 400 mpc, and for the case of the black solid bars well within a sphere of 200 mpc of radius. The crass at the center represents the position of SgrA*.

sight position is not very well constrained and the range can be quite large with respect to the radius of the sphere.

However, the result can be tested by making use of the fact the early type stars are mostly on tangential orbits and therefore show a significant anisotropy [8]. We computed the anisotropy parameter

$$\gamma = (V_T^2 - V_R^2)/(V_T^2 + V_R^2) \tag{1}$$

for 13 early type stars, covering one or more orbital time scale. For the 2 cases, 13 different random distributions of our sample in the possible ranges of line of sight positions were chosen, and the $\gamma$-value for 11 different snapshots in times calculated. Figure 6 results in the summation of 130 different combinations for the first case shown in the black solid line, and for the second case in a dotted line. The dotted histogram (200 mpc radius) shows clear anisotropy behaviour towards tangential orbits. This trend is less pronounced for the black curve (400 mpc radius).

We can conclude that the He stars of our sample are present mostly inside a sphere of about 200 mpc. A further constraint can be introduced via the presence of the mini-spiral, especially the northern arm. Here we can take into account that inspite of its presence some of the stars are not reddened (e.g IRS16 complex), and we deduce that these stars should be located in front of the mini-spiral. Vollmer & Duschl [14] give a model of the mini-spiral and describe the northern arm via gas motion within a plane and give its orientation. In Fig. 7, we show the final possible ranges in the line of sight direction obtained for our sample of stars taking al constraints into account that we mentioned above. The table gives the appropriate values of spatial positions for each star.

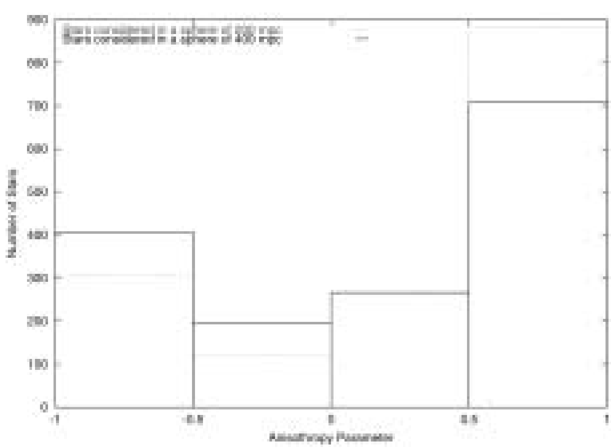


**Fig. 6.** Histogram of the anisotropy parameter $\gamma$=(${V_T}^2$-${V_R}^2$)/(${V_T}^2$+${V_R}^2$). Summation of 26 arbitrary distributions of the 13 Helium stars. The distribution covers a complete range in the line of sight positions of our sample. In dotted and black solid histogram lines, values are plotted considering that the He stars orbit inside a sphere of 200 mpc and 400 mpc of radius, respectively

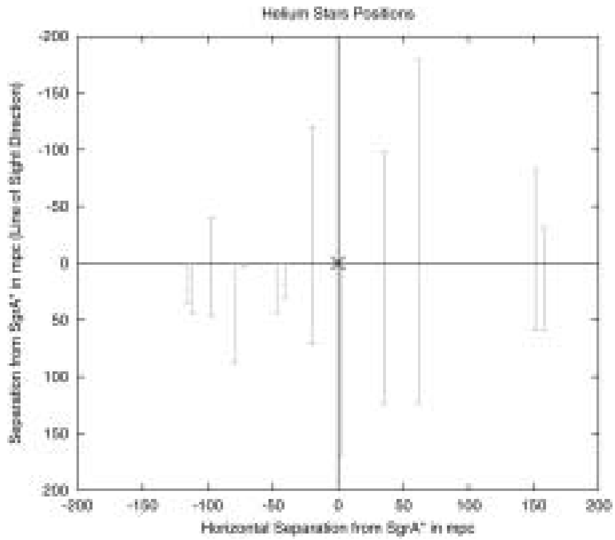



**Fig. 7.** The final possible ranges for the line of sight positions of the helium stars are represented by bars, the crass at the center represents the position of SgrA*

**Table 1.** Positions of the 13 He-stars including the ranges of possible for their minimum and maximum line of sight positions with respect to the plane of the sky that includes SgrA*.

| He Stars Names | R.A[mpc] | Dec[mpc] | Max Z[mpc] | Min Z[mpc] |
|---:|---:|---:|---:|---:|
| IRS16NW | +1.99 | +46.05 | 169 | 0.0 |
| IRS16C | -47.21 | +17.03 | 44.5 | 0.0 |
| IRS16SW | -41.00 | +37.92 | 29 | 0.0 |
| IRS16CC | -79.72 | +19.74 | 88 | 0.0 |
| IRS29N | + 61.50 | +54.57 | 123 | -180.2 |
| IRS16SE1 | -71.98 | -44.89 | 3 | 0 |
| IRS29NE1 | +35.22 | +78.95 | 124 | -98.0 |
| IRS16NE | -111.8 | +42.6 | 45 | 0.0 |
| IRS16SE2 | -114.94 | -46.4 | 35 | 0.0 |
| IRS33E | -19.35 | -127.7 | 71 | -120.0 |
| IRS7SE | -96.75 | +104.89 | 47 | -40 |
| IRS34W | +157.89 | + 62.69 | 58.5 | -31.5 |
| IRS7W | +151.32 | +193.5 | 59 | -83 |

## Acknowledgements

Support by the SFB494 at the University of Cologne and partial support by the SFB439 at the University of Heidelberg is acknowledged.

## References

1. Alexander, T., The Distribution of Stars near the Supermassive Black Hole in the Galactic Center. *The Astrophysical Journal*, **527**, Issue 2, 835-850 (1999)
2. Aarseth, S.J., From NBODY1 to NBODY6: the Growth of an Industry. *Proc. Astron. Soc. Pac*, **111**, 1333-1346
3. Baganoff, F.K., et al., Rapid X-ray flaring from the direction of the supermassive black hole at the Galactic Centre. *Nature*, **413**, 45–48 (2001)
4. Eckart, A., Genzel, R., Hofmann, R., Sams, B. J.& Tacconi-Garman, L. E.,High angular resolution spectroscopic and polarimetric imaging of the galactic center in the near-infrared. *Astrophys. J.*, **445**, L23-L26 (1995)
5. Eckart, A. & Genzel, R., Observations of stellar proper motions near the Galactic Centre. *Nature*, **383**, 415-417 (1996)
6. Eckart,A., Mouawad, N., Krips, M., Straubmeier, C., and Bertram, T., 2002, SPIE 4835-03, Cnf. Proc. of the SPIE Meeting on 'Astronomical Telescopes and Instrumentation', held in Waikoloa, Hawaii, 22-28 August 2002
7. Eckart, A., Genzel, R., Ott, T. & Schödel, R., Stellar orbits near Sagittarius A*. *Mon.Not.R.Soc.*, **331**, 917-934 (2002)
8. Genzel, R., Pichon, C., Eckart, A., Gerhard, O. & Ott, T., Stellar dynamics in the Galactic Centre: proper motions and anisotropy. *Mon.Not.R.Soc.*, **317**, 348-374 (2000)
9. Ghez, A., Morris, M., Becklin, E.E., Tanner, A. & Kremenek, T., The accelerations of stars orbiting the Milky Way's central black hole. *Nature*, **407**, 349-351 (2000)
10. Makino, J. & Aarseth, S.J., On a Hermite integrator with Ahmad-Cohen scheme for gravitational many-body problems. *Proc. Astron. Soc. Japan*, **44**, 141-151
11. Ott, T., Eckart, A., Genzel, R., Ap.J. **523**, 2480 (2000)
12. Rubilar, G. F.& Eckart, A. Periastron shifts of stellar orbits near the Galactic Center. *Astronomy and Astrophysics*, **374**, 95-104 (2001)
13. Spurzem, R., Direct N-Body Simulations. *Jl. Comp. Appl. Math*, **109**, 407-432
14. Vollmer, B. & Duschl,W.J., The Minispiral in the Galactic Center - Exploring a Data Cube With a Three Dimensional Method. *The Central Parsecs of the Galaxy, ASP Conference Series*, Ed by Heino Falcke, Angela Cotera, Wolfgang J. Duschl, Fulvio Melia, and Marcia J. Rieke, **186**, ISBN: 1-58381-012-9, p. 265 (1999)
15. Schödel, R., Ott, T., Genzel, R., Hofmann, R.; Lehnert, M.; Eckart, A., Mouawad, N., Alexander, T., Reid, M. J., Lenzen, R., Hartung, M., Lacombe, F., Rouan, D., Gendron, E., Rousset, G., Lagrange, A.-M., Brandner, W., Ageorges, N., Lidman, C., Moorwood, A. F. M., Spyromilio, J., Hubin, N., Menten, K. M.,A star in a 15.2-year orbit around the supermassive black hole at the centre of the Milky Way. *Nature*, **419**, 694-696 (2002)

# Angular Momentum Redistribution and the Evolution and Morphology of Bars

Lia Athanassoula

Observatoire de Marseille, 2 Place Le Verrier, 13248 Marseille cedex 04, France

**Abstract.** Angular momentum exchange is a driving process for the evolution of barred galaxies. Material at resonance in the bar region loses angular momentum which is taken by material at resonance in the outer disc and/or the halo. By losing angular momentum, the bar grows stronger and slows down. This evolution scenario is backed by both analytical calculations and by $N$-body simulations. The morphology of the bar also depends on the amount of angular momentum exchanged.

## 1 Introduction

Bars are common features of disc galaxies. De Vaucouleurs ([11]), using a classification based on images at optical wavelengths, found that roughly one third of all disc galaxies are barred (family SB), while yet another third have small bars or ovals (family SAB). Observations in the near infrared have shown that galaxies that had been classified as non-barred from images at optical wavelengths may have a clear bar component when observed in the near infrared. Thus Eskridge et al. ([17]) classified more than 70% of all disc galaxies as barred, while Grosbøl, Pompei & Patsis ([18]) found that only $\sim 5\%$ of all disc galaxies are definitely non barred.

Bars come in a large variety of strengths, lengths, masses, axial ratios and shapes. Great efforts have been made in order to obtain some systematics on bar structure and important advances have been made. Elmegreen & Elmegreen ([16]) have shown that earlier type bars are relatively longer (i.e. relative to the disc diameter) on average than bars in later type galaxies. They also find that early type bars have flat intensity profiles along the bar major axis, while late type bars have exponential-like profiles. A correlation has been found ([5], [27]) between the length of bars and the size of bulges. This is in good agreement with the trend found in [16], since earlier type galaxies have larger bulges than late types. Important differences between early and late type bars are also found with the Fourier decomposition of the surface density. Indeed the relative $m = 2$ and 4 components are much stronger in early than in late type bars. Moreover, the higher order components ($m = 6$ and 8), which for the late type bars are negligible, are still important for early types. Finally, the shape of the bar isodensities differ and Athanassoula et al. ([7]), using a sample of strongly barred early type galaxies, showed that their bar isophotes are rectangular-like, particularly in the region near the end of the bar.

The first trials of $N$-body simulations (e.g. [28]) show that bars grow spontaneously and are long-lived. Yet it is only recently that simulations have achieved

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 313–326, 2003.

sufficient quality to provide information on the morphology of $N$-body bars and on the mechanisms that govern bar formation and bar evolution. I will here discuss some of the latest results of $N$-body simulations. I will argue that it is the exchange of angular momentum within the galaxy that will determine the bar strength and the rate at which the pattern speed decreases after the bar has formed, as well as the bar morphology.

## 2 Angular Momentum Exchange and Bar Evolution

Exchange of energy and angular momentum between stars at resonance with a spiral density wave has been first discussed by Lynden-Bell & Kalnajs ([26]). Using linear perturbation theory, these authors showed that, for a steady forcing, stars emit, or absorb, angular momentum only if they are at resonance. Stars at the inner Lindblad resonance (hereafter ILR) lose angular momentum, while stars at the outer Lindblad resonance (hereafter OLR) gain it. This groundbraking work has to be extended in order to be applied to bars in general and $N$-body bars in particular. HI observations, basically starting with [10], have now established that, if Newton's law of gravity is valid, then disc galaxies are embedded in a dark matter component, called the halo, whose mass exceeds that of the disc. This component should now be taken into account as an extra partner in the angular momentum exchange process. Furthermore, bars are strongly non-linear features, since they contain a large fraction of the mass in the inner parts of the disc and a considerable part of the total disc mass. Thus any linear theory should be thought of as a guiding line, to be supplemented by and tested against adequate $N$-body simulations. It is obvious that such simulations should be fully self-consistent, since rigid components can not exchange energy and angular momentum.

In [4] I extended the analytical work of [26] to include spheroidal components, like a halo and/or a bulge, and also supplemented it with fully self-consistent $N$-body simulations. In the analytical part I showed that, if the distribution function of the spheroidal component is a function only of the energy, then at all resonances the halo and bulge particles can only gain angular momentum. Also, since the bar is a strongly nonlinear feature, higher multiplicity resonances should be taken into account. Thus angular momentum is emitted by particles (stars) at the resonances in the inner disc, mainly the ILR, but also the inner -1:m resonances nearer to corotation (hereafter CR). It is absorbed by disc particles (stars) at the OLR, or the outer 1:m resonances, outside corotation, or by the resonant particles in the halo and/or bulge components. Since the bar is a negative angular momentum perturbation ([26]), by losing angular momentum it will grow. This clearly outlines a scenario for the evolution of barred galaxies.

## 3 The Effect of the Halo on Bar Evolution

As the bar loses angular momentum, it grows stronger. This, however, can only happen if there are absorbers that can absorb the angular momentum that the bar region emits. Thus the existence of a massive halo component, whose resonances can absorb considerable amounts of angular momentum, will help the bar grow. At first sight this may be thought to go against old claims that haloes stabilise bars. In fact, a more precise wording is necessary. Indeed, the halo slows down the bar growth in the initial stages of the evolution. At later stages, however, the situation can be reversed, since the halo may absorb the angular momentum emitted by the bar, and thus it may allow the latter to grow further. Thus bars that grow in halo-dominated discs can be stronger than bars that grow in disc-dominated surroundings. This effect was not seen till recently, since the older studies were either 2D (e.g. [33], [8]), or 3D but with few particles (e.g. [31]), or had rigid haloes. In all these cases the halo was denied from the onset its destabilising influence. Its effect becomes clear in fully self-consistent $N$-body simulations, with an adequate particle number and resolution. Thus [6] showed that stronger bars can grow in cases with more important halo components.

The influence of the halo is also illustrated in Fig. 1, where I compare the results of two $N$-body simulations. Initially their disc is exponential, with unit mass and scale-length ($M_d = 1$, $R_d = 1$) and its $Q$ parameter ([32]) is independent of radius and equal to 1.2. Since $G = 1$, taking the mass of the disc equal to $5 \times 10^{10}$ $M_{\odot}$, and its scale-length equal to 3.5 kpc implies that the unit of velocity is 248 km/sec and the unit of time is $1.4 \times 10^7$ yrs. This calibration is reasonable, but is not unique, so in the following I will give all quantities in computer units. The reader can then convert the values to astronomical units according to his/her needs. The halo component is spherical, non-rotating and has an isotropic velocity distribution. It follows a pseudo-isothermal radial density profile ([19]) and has a total mass $M_h = 5$, a core radius $\gamma = 0.5$ and a cutoff radius of $r_c = 10$. Its density is truncated at 15 disc scale-lengths, i.e. at a radius containing more than 96% of its mass. In building the initial conditions I loosely followed [19] and [6], and the simulations were run on the Marseille GRAPE-5 systems (for a description of the GRAPE-5 boards see [22]). The only difference between the initial conditions of the two simulations is that for simulation LH, illustrated in the left panels, the halo is live and represented by roughly $10^6$ particles, while for simulation RH, illustrated in the right panels, it is rigid, i.e. represented by an analytical potential and thus can neither emit nor absorb angular momentum. Although their initial conditions are so similar, the two simulations evolve in a very different way. After some initial multi-spiral episodes, LH forms a bar which grows stronger with time. Its morphology at $t = 700$ is shown in the left panels. The bar is long and strong and has ansae-type features near the end of its major axis. It is surrounded by a ring, which can be compared to the inner rings often observed in barred galaxies. The bar formation entails considerable redistribution of the disc matter, both radially and azimuthally. On the other hand the disc in simulation RH stays close to axisymmetric, except for some multi-armed spirals which die out with time. Only at the latest stages of the evolution does it form an oval distortion, and even that is weak and short and is confined to the innermost parts of the disc, as can be seen for $t = 900$ in the right

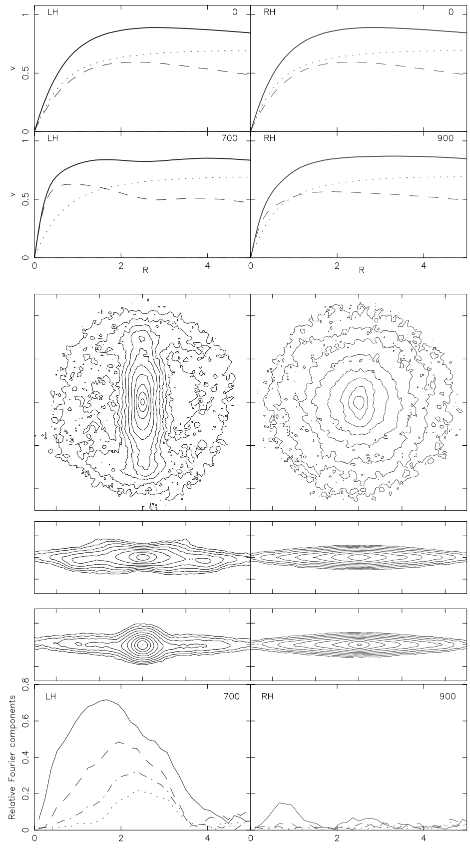



**Fig. 1.** Basic information on simulations LH (time $t$ = 600) and RH (time $t$ = 900). The two upper rows give the circular velocity curves at time 0 and $t$. The dashed and dotted lines give the contributions of the disc and halo respectively, while the thick full lines give the total circular velocity curves. The third row of panels gives the isocontours of the density of the disc particles projected face-on and the fourth and fifth row give the side-on and end-on edge-on views, respectively. The side of the box for the face-on views is 10 length units and the height of the box for the edge-on views is 3.33. The isodensities in the third row of panels have been chosen so as to show best the features in the bar and in the inner disc. No isodensities for the outer disc have been included, although the disc extends beyond the area shown in the figure. The sixth row of panels gives the $m$ = 2, 4, 6, and 8 Fourier components of the mass.

panels of Fig. 1. I show this simulation at a later time than that for simulation LH because at earlier times there is very little structure visible.

Seen edge-on with the bar seen side-on (i.e. with the line of sight along the bar minor axis), simulation LH exhibits a very strong peanut, which is totally absent from simulation RH (fourth row of panels). Seen edge-on with the bar seen end-on (i.e. with the line of sight along the bar major axis), the peanut in LH resembles a bulge (left panel on fifth row). This underlines the hazards involved in classifying edge-on galaxies, since the classifier may in such cases easily misinterpret the bar for a bulge.

A useful way of quantifying the bar strength is with the help of the Fourier components of the mass, or density. These can be defined as

$$F_m(r) = \sqrt{A_m^2(r) + B_m^2(r)}/A_0(r), \qquad m = 0, 1, 2, ..... \tag{1}$$

where

$$A_m(r) = \frac{1}{\pi}\int_0^{2\pi} \Sigma(r,\theta) cos(m\theta) d\theta, \qquad m = 0, 1, 2, ..... \tag{2}$$

and

$$B_m(r) = \frac{1}{\pi}\int_0^{2\pi} \Sigma(r,\theta) sin(m\theta) d\theta, \qquad m = 1, 2, ..... \tag{3}$$

For runs LH and RH, these components for $m = 2$, 4, 6 and 8 are shown in the lower panels of Fig. 1. For run LH all four components are important, due to the strength of the bar. Their amplitude decreases with increasing $m$, while the location of the maximum shifts outwards. On the other hand, for model RH only the $m = 2$ component stands out from the noise, but its amplitude is rather small, smaller than e.g. that of the $m = 8$ for model LH.

Since the only difference between the initial conditions of models LH and RH is that the halo of the one is responsive, while that of the other is rigid, we can conclude that the halo response is crucial for determining the evolution of barred galaxies.

## 4 Bar Slow-Down

I ran a large number of simulations similar to those described in the previous sections. I noted that, as it loses angular momentum, the bar grows longer, and more massive, thus stronger. Angular momentum loss, however, is not only linked to an increase in the bar strength. It is also linked to a slow-down, i.e. to a decrease of the bar pattern speed $\Omega_p$ with time. Such a slow-down has indeed been seen in a number of simulations and has also been predicted analytically ([34], [36], [24], [25], [20], [1], [12], [13]). It can also be seen in Fig. 2, which shows the run of the bar pattern speed with time for simulation LH, whose morphology at time $t = 700$ is shown in the left panels of Fig. 1. Note that the bar slows down considerably with time.

## 5 Resonances

In order for haloes to be able to absorb angular momentum, they need to have a considerable fraction of their mass at resonance. This was shown to be true in ([2]). I will illustrate it here for model LH. The procedure is the same as that followed in [2]. I calculate the potential from the mass distribution in the

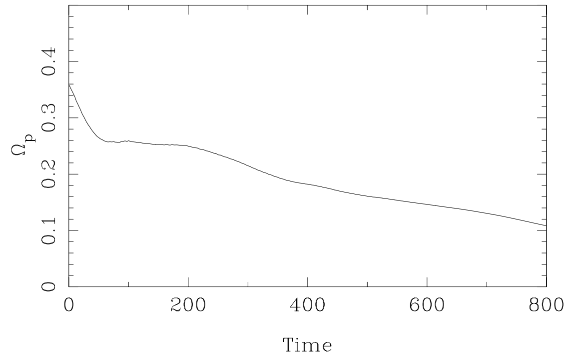



**Fig. 2.** Bar pattern speed of simulation LH as a function of time.

disc and halo component at time $t = 800$, by freezing all motion except for the bar, to which I assign bulk rotation with a pattern speed equal to that found in the simulation at that time. I then pick at random 100 000 disc and 100 000 halo particles and, using their positions and velocities as initial conditions, I calculate their orbits for 40 bar rotations. Using spectral analysis ([9], [23]), I then find the principal frequencies of these orbits, i.e. the angular velocity $\Omega$, the epicyclic frequency $\kappa$ and the vertical frequency $\kappa_z$. An orbit is resonant if there are three integers $l$, $m$ and $n$, such that

$$l\kappa + m\Omega + n\kappa_z = -\omega_R = m\Omega_p \tag{4}$$

Orbits on planar resonances fulfill

$$l\kappa + m\Omega = -\omega_R = m\Omega_p \tag{5}$$

The ILR corresponds to $l = -1$ and $m = 2$, CR to $l = 0$, and OLR to $l = 1$ and $m = 2$.

The upper panels of Fig. 3 show, for time $t = 800$, the mass per unit frequency ratio $M_R$ of particles having a given value of the frequency ratio $(\Omega - \Omega_p)/\kappa$ as a function of this frequency ratio[1]. The distribution is not uniform, but has clear peaks at the location of the main resonances, as was already shown in [2] and [4]. The highest peak for the disc component is at the ILR, followed by $(-1, 6)$ and CR. In all simulations with strong bars the ILR peak is strong. The existence of peaks at other resonances as well as their importance varies from one run to another and also during the evolution of a given run. For example the CR peak is, in many other simulations, much stronger than in the example shown here. For the spheroidal component the highest peak is at CR, followed by peaks at the ILR, OLR and $(-1, 4)$.

The lower panels show the angular momentum exchanged. For this I calculated the angular momentum of each particle at time 800 and at time 500, as described in [4], and plotted the difference as a function of the frequency ratio

[1] See [4] for more information on $M_R$ and on how it is derived.

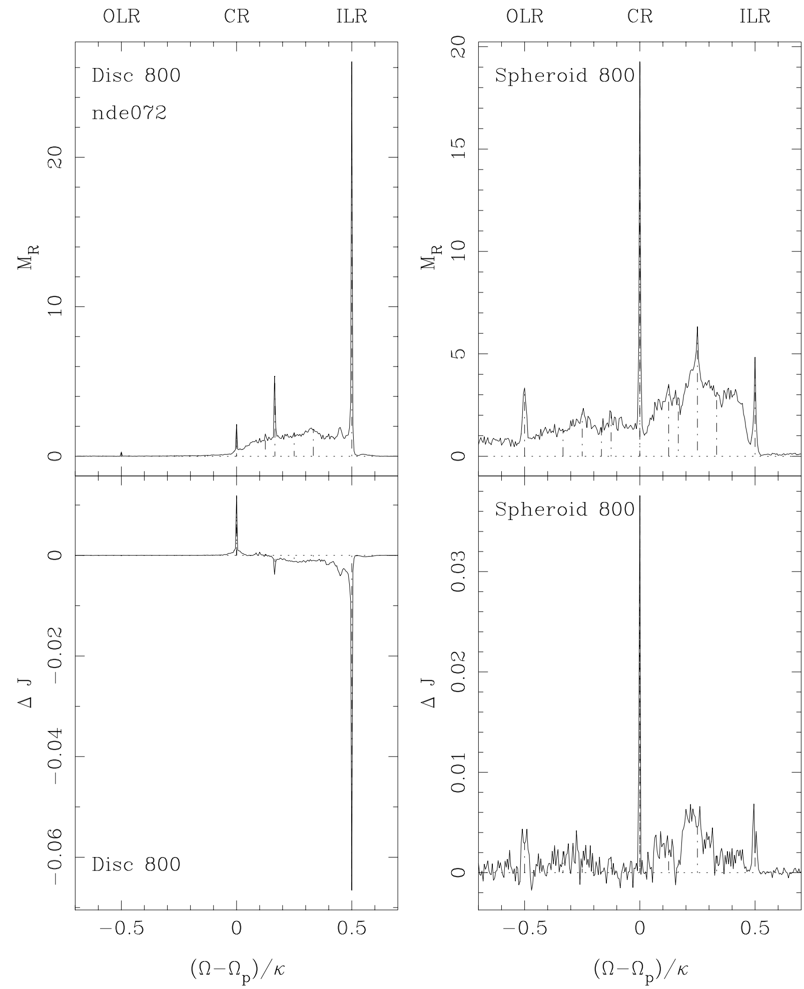



**Fig. 3.** The upper panels give, for time $t = 800$, the mass per unit frequency ratio, $M_R$, as a function of that ratio. The lower panels give $\Delta J$, the angular momentum gained or lost by the particles between times 800 and 500, plotted as a function of their frequency ratio $(\Omega - \Omega_p)/\kappa$, calculated at time $t = 800$. The left panels correspond to the disc component and the right ones to the halo. The component and the time are written in the upper left corner of each panel. The vertical dot-dashed lines give the positions of the main resonances.

of the particle at time 800. It is clear from the figure that disc particles at ILR and at the $(-1, 6)$ resonance lose angular momentum, while those at CR gain it. There is a also a general, albeit small, loss of angular momentum from particles with frequencies between CR and ILR. This could be partly due to particles trapped around secondary resonances, and partly due to angular momentum taken from particles which are neither resonant, nor near-resonant, but can still lose a small amount of angular momentum because the bar is growing. The corresponding panel for the spheroid is, as expected, more noisy, but shows that particles at all resonances gain angular momentum. Thus this plot, and similar ones which I did for other simulations, confirm the analytical results of [4], and show that the linear results concerning the angular momentum gain or loss by resonant particles, qualitatively at least, carry over to the strongly nonlinear regime.

## 6 What Determines the Strength of Bars and Their Slow-Down Rate?

I have shown in the previous sections that the halo can take angular momentum from the bar, thus making it stronger and slower. However, for this effect to be important, the amount of angular momentum exchanged must be considerable. For the latter to happen the halo must

- be sufficiently massive in the regions containing the principal resonances.
- not be too hot, i.e. not have too high velocity dispersion. Indeed, hot haloes can not absorb much angular momentum, even if they are massive (e.g. [4]).

Thus the length and the slow-down rate of bars are naturally limited by the mass and velocity distribution of the halo. Examples of this can be found in [4].

## 7 Trends and Correlations

In [4] I found trends and correlations between the angular momentum absorbed by the spheroid (i.e. the halo plus, whenever existent, the bulge), the bar strength and the bar pattern speed. They are based on a set of simulations analogous to those described in the previous sections. Such plots are given also in Fig. 4, based on a somewhat larger sample of simulations. About three quarters of them were run with the Marseille GRAPE-5 systems, and roughly one quarter was run on PCs using Dehnen's treecode ([14], [15]). Each point represents one simulation and the trends are the same as those found in [4]. The upper panels show the results for the whole sample, the middle panels contain only simulations where the halo has a small core radius ($\gamma < 2$), $M_h = 5$ and does not extended beyond 15 disc scale-lengths, and the lower ones contain only simulations where the halo has a large core radius ($\gamma > 2$), $M_h = 5$ and again does not extended beyond 15 disc scale-lengths.

The right panel shows that there is a correlation between the angular momentum of the spheroid and the bar strength. This correlation holds also when I restrict myself to simulations with large (or small) core radii as seen in the second and third row. A trend also exists between the spheroid angular momentum and the bar pattern speed. In this case, however, simulations with large core radii behave differently from those with small radii. Indeed, for simulations with a small core radius (i.e. centrally concentrated halos) I find a very strong correlation

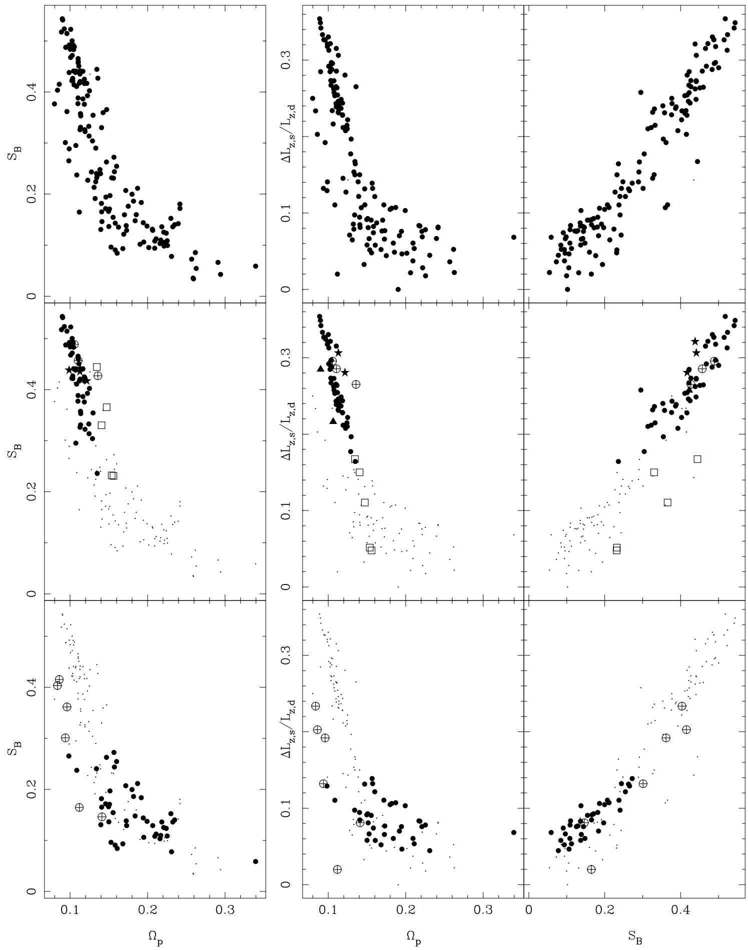



**Fig. 4.** Relations between the bar strength and the pattern speed (left panels), the spheroid angular momentum and the pattern speed (middle panels) and the spheroid angular momentum and the bar strength (right panels), at times $t = 800$. The spheroid angular momentum is normalised by the initial disc angular momentum ($L_{z,d}$). The simulations under consideration in each panel are marked with a filled circle and the rest by a dot. The upper row includes all simulations, the middle and the lower ones subsamples, as described in the text. In the middle panel simulations with a bulge are marked with a $\oplus$, simulations with $\gamma = 0.01$ with a filled star, simulations with $2 > \gamma \geq 1$ with a filled triangle and simulations with $Q_{init} \geq 2$ with an open square. In the lower panel simulations with $Q_{init} \geq 1.4$ and $z_0 \geq 0.2$ are marked with an $\oplus$.

between the spheroid angular momentum and the pattern speed, particularly if I restrict myself to one value of $\gamma$. In such simulations the angular momentum is exchanged primarily between the bar region and the spheroid, thus accounting for the very tight correlation. Simulations with large cores behave differently (lower middle panel). They show only a rough trend, except for simulations with a hot disc, which show a tight correlation. This is easily explained in the scenario of evolution via angular momentum exchange. The outer parts of hot discs absorb only little angular momentum, so that the exchange is basically between the bar region and the spheroid, thus accounting for the tight correlation. On the other hand, if the outer disc is cold, then it can participate more actively in the exchange. Since the angular momentum absorbed by the spheroid (plotted in Fig. 4) is not the total angular momentum exchanged, but only a fraction of it, I find only a trend.

## 8 Comparing the Morphology of *N*-Body and of Real Bars

The correlations discussed in section 7 show clearly that models that have exchanged more angular momentum have stronger bars than models that have exchanged little. By examining the results of the individual simulations, I could see that, in cases where large amounts of angular momentum have been exchanged, the bars are long, relatively thin and have rectangular-like isodensities, particularly in their outer parts. A typical example of such a case is given in the left panels of Fig. 5 (see also [6]). Note also the existence of ansae at the ends of the bar, a feature sometimes observed in early type barred galaxies. On the other hand, models that have exchanged little angular momentum have less homogeneous properties. For example they can have either ovals, or short bars. Typical examples of such cases are given in the middle and right panels of Fig. 5, respectively. The model in the left panel has exchanged about 15 percent of the disc angular momentum by the time shown in Fig. 5, while the other two models only of the order of a percent.

The edge-on morphology also is strongly influenced by the amount of angular momentum exchanged. The strong bar, when seen edge-on, displays a clear peanut morphology, as often observed. On the other hand the oval has a boxy edge-on appearance, while the small bar has not changed significantly the edge-on morphology of the galaxy.

The difference in bar strength is also illustrated in the lower panels of Fig. 5, which show the relative Fourier components of the density for $m = 2$, 4, 6 and 8 for the three simulations. The simulation that exchanged a lot of angular momentum has a very strong $m = 2$ component, with a secondary maximum roughly at the position of the ansae. The remaining components, even the $m = 6$ and 8 ones, are also important. The location of their maximum moves outwards with increasing $m$. The oval has much lower Fourier components, and only the $m = 2$ stands out from the noise. For small radii all components are nearly zero, which means that the oval must be nearly axisymmetric in its innermost parts. On the other hand, the $m = 2$ amplitude drops slowly with radius in the outer parts, thus extending to large radii. The small bar has Fourier components which drop rapidly with radius, i.e. they are noticeable only in the central region, as expected since the bar is confined there. Only the $m = 2$ and 4 components stand out from the noise.

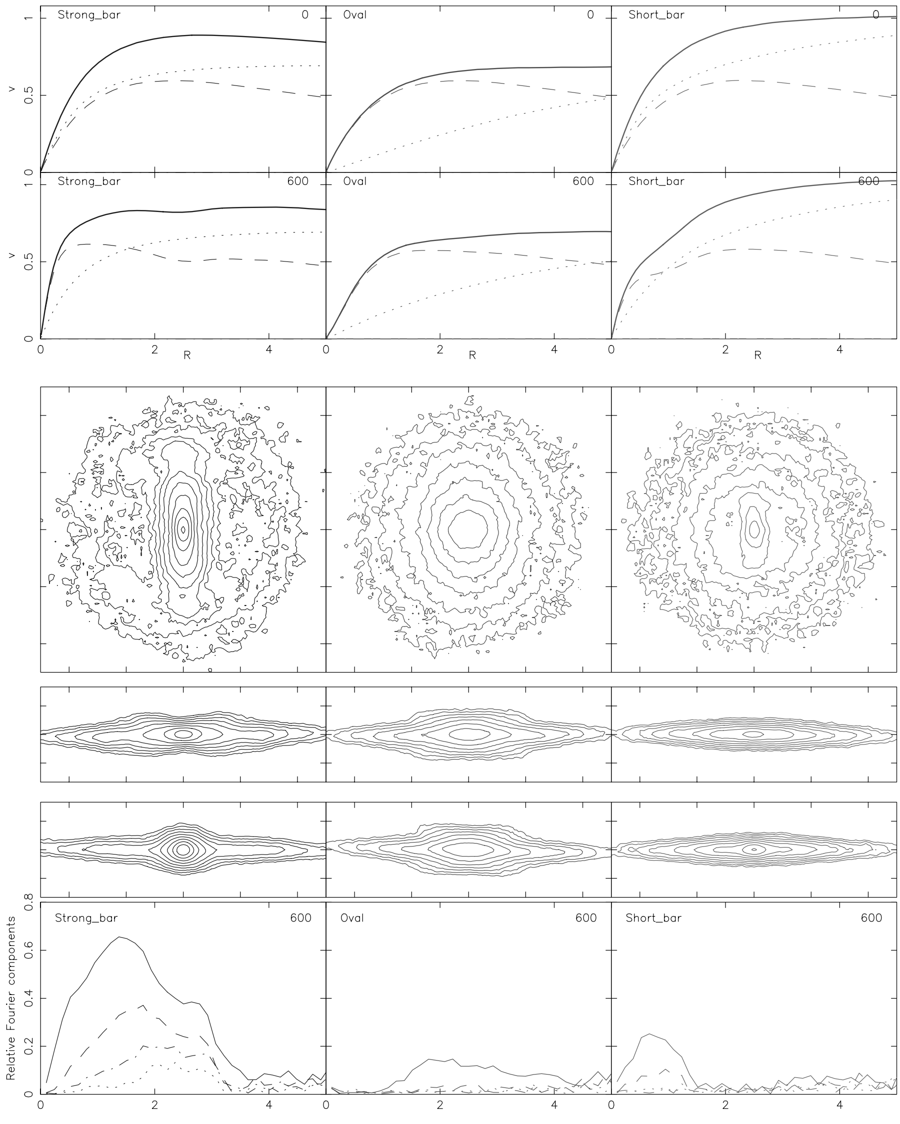



**Fig. 5.** Comparison of a simulation forming a strong bar (left panels), one forming an oval (middle panels) and one forming a short bar (right panels). The layout is as for Fig. 1.

The radial rearrangement of the disc material due to the bar can be inferred by comparing the initial with the current circular velocity curves, given in the first and second rows of panels. The strong bar has entailed a substantial radial rearrangement, the final disc mass distribution being considerably more centrally concentrated than the initial one. On the other hand, in the other two simula-

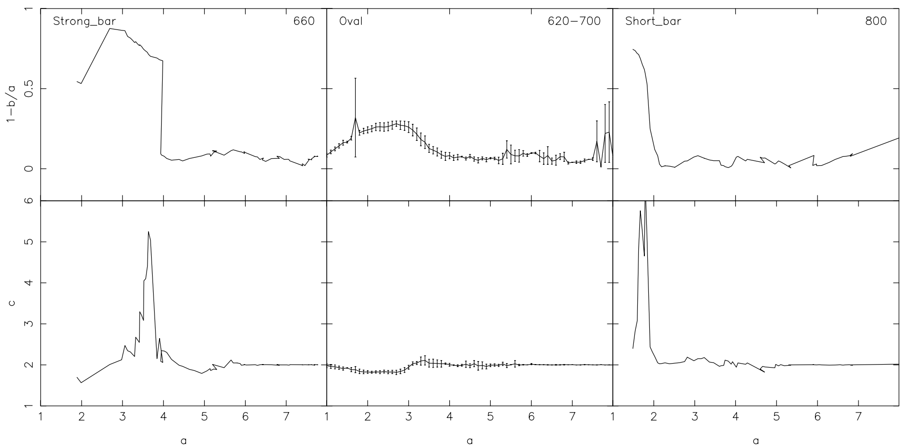



**Fig. 6.** The upper panels show the run of the ellipticity 1 - $b/a$ as a function of the semi-major axis $a$. The lower panels show the run of the shape parameter $c$, also as a function of $a$. The left panels corresponds to a model with a strong bar, the middle ones to model with an oval and the right ones to a model with a short bar. To improve the signal-to-noise ratio for the model with the oval I took an average over a time interval, namely [620-700]. The dispersion during that time is indicated by the error bars. The times are given in the upper right corner of the upper panels.

tions, and particularly in the one producing the oval, there is very little radial rearrangement of the disc material. Since there is also hardly any radial rearrangement of the halo material ([6], [3], [35]), this means that the disc-to-halo mass ratio changes most in the simulations where more angular momentum has been exchanged.

Quantitative comparison of the bar form of the three models is given in Fig. 6. The values of the bar semi-major and semi-minor axes ($a$ and $b$, respectively) and of the shape parameter ($c$) were obtained by fitting generalised ellipses of the form

$$(|x|/a)^c + (|y|/b)^c = 1, \tag{6}$$

to the bar isodensities. The shape parameter $c$ is 2 for ellipses, larger than 2 for rectangular-like generalised ellipses, and smaller than 2 for diamond-like ones. From this figure one can note that both the strong and the short bar are thin, and in general see how their axial ratios vary with the semi-major axis. The shape parameter is given in the lower panels. We see that both the strong and the short bar have rectangular-like isodensities in the outer regions of the bar, while the oval has a shape very close to elliptical. In fact the strong bar has axial ratios and shapes very similar to those found in [7] by applying the same type of analysis to a sample of early type barred galaxies.

Plotting the run of the density along the bar major axis ([6]) I find for the strong bar a profile which is rather flat within the bar region, similar to what was found in [16] for early type bars.

It is thus clear that the amount of angular momentum exchanged influences the morphology of the bar. In my first example, where a lot of angular momentum was transferred from the bar to the outer halo (mainly), the result is a long, strong bar, with some rectangular-like isophotes and ansae at its ends. The examples where little angular momentum was exchanged have a very different morphology, one forming an oval and the other a short bar. What determines which one of the two it will be? In the examples shown here, and in a rather large sample of similar cases, the oval was formed in an initially hot disc, while the short bar grew in a hot halo. The existing theoretical framework, however, gives no predictions on this point and work is in progress to elucidate this further.

## 9 Summary

In this paper I reviewed evidence that shows that angular momentum is exchanged between the bar region in the one hand, and the outer disc and the spheroid on the other. This exchange determines the slow-down rate of the bar, as well as its strength and its overall morphology.

### Acknowledgments

I thank the organisers for inviting me to this interesting conference. I thank M. Tagger, A. Bosma, W. Dehnen, A. Misiriotis, C. Heller, I. Shlosman, F. Masset, J. Sellwood, O. Valenzuela, A. Klypin and P. Teuben for many stimulating discussions. I thank J. C. Lambert for his help with the GRAPE software and with the administration of the simulations and W. Dehnen for making available to me his tree code and related programs. I also thank the INSU/CNRS, the University of Aix-Marseille I, the Region PACA and the IGRAP for funds to develop the GRAPE and Beowulf computing facilities used for the simulations discussed in this paper and for their analysis.

## References

1. E. Athanassoula: ‘Evolution of bars in isolated and in interacting disk galaxies’. In *Barred Galaxies*, eds. R. Buta, D. Crocker and B. Elmegreen (Astron. Soc. of the Pacific Conference series, 91), pp. 309–321 (1996)
2. E. Athanassoula: Astrophys. J. **569**, L83 (2002)
3. E. Athanassoula: ‘Formation and Evolution of Bars in Disc Galaxies’. In *Disks of Galaxies: Kinematics, Dynamics and Perturbations*, eds. E. Athanassoula, A. Bosma, R. Mujica (Astron. Soc. of the Pacific Conference Series, 275) pp. 141–152 (2002)
4. E. Athanassoula: Mon. Not. R. Astron. Soc., in press (2003)
5. E. Athanassoula, L. Martinet: Astron. Astrophys. **87**, L10 (1980)
6. E. Athanassoula, A. Misiriotis: Mon. Not. R. Astron. Soc. **330**, 35 (2002)
7. E. Athanassoula, S. Morin, H. Wozniak, et al.: Mon. Not. R. Astron. Soc. **245**, 130 (1990)
8. E. Athanassoula, J. A. Sellwood: Mon. Not. R. Astron. Soc. **221**, 213 (1986)

9. J. Binney, D. Spergel: Astrophys. J. **252**, 308 (1982)
10. A. Bosma: Astron. J., **86**, 1825 (1981)
11. G. de Vaucouleurs: Astrophys. J. Suppl. **8**, 31 (1963)
12. V. P. Debattista, J. A. Sellwood: Astrophys. J. **493**, L5 (1988)
13. V. P. Debattista, J. A. Sellwood: Astrophys. J. **543**, 704 (2000)
14. W. Dehnen: Astrophys. J. **536**, L39 (2000)
15. W. Dehnen: J. Comp. Phys. **179**, 27 (2002)
16. B. G. Elmegreen, D. M. Elmegreen: Astrophys. J. **288**, 438 (1985)
17. P. B. Eskridge, J. A. Frogel, R. W. Pogge, et al.: Astron. J., **119**, 536 (2000)
18. P. Grosbøl, E. Pompei, P. Patsis: 'Spiral Structure Observed in Near-Infrared'. In *Disks of Galaxies: Kinematics, Dynamics and Perturbations*, eds. E. Athanassoula, A. Bosma, R. Mujica (Astron. Soc. of the Pacific Conference Series, 275) pp. 305–310 (2002)
19. L. Hernquist: Astrophys. J. Suppl. **86**, 389 (1993)
20. L. Hernquist, M. D. Weinberg: Astrophys. J. **400**, 80 (1992)
10. F. Hohl: Astrophys. J. **168**, 343 (1971)
22. A. Kawai, T. Fukushige, J. Makino, M. Taiji: Pub. Astron. Soc. Japan **152**, 659 (2000)
23. J. Laskar: Icarus, **88**, 266 (1990)
24. B. Little, R. G. Carlberg: Mon. Not. R. Astron. Soc. **250**, 161 (1991a)
25. B. Little, R. G. Carlberg: Mon. Not. R. Astron. Soc. **251**, 227 (1991b)
26. D. Lynden-Bell, A. J. Kalnajs: Mon. Not. R. Astron. Soc. **250**, 161 (1972)
27. P. Martin: Astron. J. **109**, 2428 (1995)
28. R. H. Miller, K. H. Prendergast, W. J. Quirk: Astrophys. J. **161**, 903 (1979)
29. K. Ohta: 'Global Photometric Properties of Barred Galaxies'. In *Barred Galaxies*, eds. R. Buta, D. A. Crocker, B. G. Elmegreen (Astron. Soc. of the Pacific Conference Series, 91) pp. 37–43, (1996)
30. K. Ohta, M. Hamabe, K. Wakamatsu: Astrophys. J. **357**, 71 (1990)
31. J. P. Ostriker, P. J. E. Peebles: Astrophys. J. **186**, 467 (1973)
32. A. Toomre: Astrophys. J. **139**, 1217 (1964)
33. A. Toomre: 'What amplifies the spirals' In: *The Structure and Evolution of Normal Galaxies*, ed. S. M. Fall, D. Lynden-Bell (Cambridge University Press) pp. 111–136 (1981)
34. S. Tremaine, M. D. Weinberg: Mon. Not. R. Astron. Soc. **209**, 729 (1984)
35. O. Valenzuela, A. Klypin: in preparation (2003)
36. M. D. Weinberg: Mon. Not. R. Astron. Soc. **213**, 451 (1985)

# Major Mergers and the Origin of Elliptical Galaxies

Andreas Burkert[1] and Thorsten Naab[2]

[1] Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany
[2] Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK

**Abstract.** The formation of elliptical galaxies as a result of the merging of spiral galaxies is discussed. We analyse a large set of numerical N-Body merger simulations which show that major mergers can in principle explain the observed isophotal fine structure of ellipticals and its correlation with kinematical properties. Equal-mass mergers lead to boxy, slowly rotating systems, unequal-mass mergers produce fast rotating and disky ellipticals. However, several problems remain. Anisotropic equal mass mergers appear under certain projections disky which is not observed. The intrinsic ellipticities of remnants are often larger than observed. Finally, although unequal-mass mergers produce fast rotating ellipticals, the remnants are in general more anisotropic than expected from observations. Additional processes seem to play an important role which are not included in dissipationless mergers. They might provide interesting new information on the structure and gas content of the progenitors of early-type galaxies.

## 1 Introduction

Giant elliptical galaxies are believed to be very old stellar systems that formed by a major merger event preferentially very early at a high redshift of more than two ([59], [61]). The merger triggered an intensive star-formation phase which turned most of the gas of the progenitors into stars. Some fraction of the gas was heated to temperatures of order the virial temperature, producing X-ray coronae which are still visible today. The stellar disks of the progenitors were destroyed as a result of the strong tidal forces during the merger, leading to kinematically hot, spheroidal stellar remnants. Subsequently, the systems experienced very little accretion and merging with negligible star formation [22]. This scenario is supported by many observations which indicate that ellipticals contain stellar populations that are compatible with purely passive evolution ([21], [1], [27], [65]). or with models of an exponentially, fast decreasing star formation rate [66].

An alternative scenario which is based on hierarchical theories of galaxy formation predicts that massive galaxies are assembled relatively late in many generations of mergers through multiple mergers of small subunits, with additional smooth accretion of gas ([38], [39]). In this case, ellipticals might form either if the multiple subunits are already preferentially stellar or if star formation was very efficient during the protogalactic collapse phase [42].

The idea that ellipticals form from major mergers of massive disk galaxies has been originally proposed by Toomre & Toomre [61]. Their "merger hypothesis"

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 327–339, 2003.

has been explored in details by many authors, using numerical simulations. Gerhard [31], Negroponte & White [52], Barnes [2] and Hernquist [33] performed the first fully self-consistent merger models of two equal-mass stellar disks embedded in dark matter halos. The remnants are slowly rotating, pressure supported and anisotropic. They generally follow an $r^{1/4}$ surface density profile for radii $r \geq 0.5 r_e$, where $r_e$ is the effective radius. However it turns out that due to phase space limitations [23], an additional massive central bulge component is required [35], to fit the observed de Vaucouleurs profile [19] also in the inner regions. All simulations demonstrated consistently that the global properties of equal mass merger remnants resemble those of ordinary slowly rotating massive elliptical galaxies.

More recently it has become clear that ellipticals have quite a variety of fine structures with peculiar kinematical properties which, in contrast to their universal global properties, can give a more detailed insight into their formation history. It is interesting to investigate whether the merging hypothesis can explain these observations and, if yes, whether they provide more information on the validity of this scenario, the orbital parameters of the mergers and the structure and gas content of the progenitors from which the ellipticals formed.

Elliptical galaxies can be subdivided into two major groups with respect to their structural properties ([9], [7], [8], [41]). Faint ellipticals are isotropic rotators with small minor axis rotation and disky deviations of their isophotal shapes from perfect ellipses. Their isophotes are peaked in the rotational plane and a Fourier analyses of the isophotal deviation from a perfect ellipse leads to a positive value of the fourth order coefficient $a_4$. These galaxies might contain secondary, faint disk components which contribute up to 30% to the total light in the galaxy, indicating disk-to-bulge ratios that overlap with those of S0-galaxies ([55], [58]). Disky ellipticals have power-law inner density profiles ([43], [28]) and show little or no radio and X-ray emission [10]. Most massive ellipticals have boxy isophotes, with negative values of $a_4$. They also show flat cores ([43], [28] Faber et al. 1997) and their kinematics is more complex than that of disky ellipticals. Boxy ellipticals rotate slowly, are supported by velocity anisotropy and have a large amount of minor axis rotation. Like the secondary disks of disky ellipticals, the boxy systems occasionally reveal kinematically decoupled core components, that most likely formed from gas that dissipated its orbital energy during the merger, accumulated in the center and subsequently turned into stars ([29], [57], [8]). The cores inhibit flattened rapidly rotating disk- or torus-like stellar structures that dominate the light in the central few hundred parsecs ([56], [45]), but they contribute only a few percent to the total light of the galaxy. The fact that the stars are metal-enhanced confirms that gas infall and subsequently violent star formation, coupled with metal-enrichment must have played an important role in the centers of merger remnants ([11], [25], [12], [26]). Boxy ellipticals show strong radio emission and high X-ray luminosities, resulting from emission from hot gaseous halos [16] that probably formed from gas heating during the merger. These hot gaseous bubbles are however absent in disky ellipticals. The distinct physical properties of disky and boxy elliptical galaxies indicate that both types of ellipticals experienced different formation histories.

In order to understand the origin of boxy and disky ellipticals the isophotal shapes of the numerical merger remnants have been investigated in detail. It has been shown that the same remnant can appear either disky or boxy when viewed from different directions (Hernquist 1993b) with a trend towards boxy isophotes ([36], [58]). Barnes [4] and Bendo & Barnes [15] analysed a sample of disk-disk mergers with a mass ratio of 3:1 and found that the remnants are flattened and fast rotating in contrast to equal mass mergers. Naab et al. [47] studied the photometrical and kinematical properties of a typical 1:1 and 3:1 merger remnant in details and compared the results with observational data . They found an excellent agreement and proposed that fast rotating disky elliptical galaxies can originate from purely collisionless 3:1 mergers while slowly rotating, pressure supported ellipticals form from equal mass mergers of disk galaxies.

Despite these encouraging results no systematic high-resolution survey of mergers has yet been performed to explore the parameter space of initial conditions and specify the variety of properties of merger remnants that could arise. Recently, Naab & Burkert [50] completed a large number of 112 merger simulations of disk galaxies adopting a statistically unbiased sample of orbital initial conditions with mass ratios $\eta$ of 1:1, 2:1, 3:1, and 4:1. This large sample allows a much more thorough investigation of the statistical properties of merger remnants in comparison with observed disky and boxy ellipticals.

## 2 The Merger Models

Cosmological simulations currently are not sophisticated enough to predict initial conditions of major spiral mergers. Some insight can however be gained by investigating the typical conditions under which dark matter halos merge in standard cold dark matter models. Such a detailed analysis was done by Khochfar, Burkert & White [40]. The first encounter is in most cases a parabolic orbit with an impact parameter of order the scale radius of the more massive dark halo, with random orientation of the net spin axes of the progenitors. Unequal mass mergers with mass ratios $\eta$ of 3:1 to 4:1 are as likely as equal-mass mergers with $\eta = 1:1-2:1$. The cold dark matter simulations however do not provide information on the internal structure and gas content of the merging spirals. In fact, simulations of hierarchical structure formation including gas lead to disk galaxies which do not fit the zero point of the Tully-Fisher relation with disk scale radii that are up to a factor of 10 smaller than observed [51]. Unless these problems are solved we cannot study the subsequent merging of disk galaxies self-consistently, including the large-scale evolution of the Universe. In the meantime, the best strategy is to construct plausible equilibrium models of disk galaxies and investigate their merging in isolation.

Equilibrium spirals were generated using the method described by Hernquist [34]. The following units are adopted: gravitational constant G=1, exponential scale length of the larger disk $h = 1$ and mass of the larger disk $M_d = 1$. For a typical spiral like the Milky Way these units correspond to $M_d = 5.6 \times 10^{10} M_\odot$, h=3.5 kpc and a unit time of $1.3 \times 10^7$ yrs. Each galaxy consists of an exponential disk, a spherical, non-rotating bulge with mass $M_b = 1/3$ and a Hernquist density profile [32] with a scale length $r_b = 0.2$. The stellar system is embedded in a spherical pseudo-isothermal halo with a mass $M_d = 5.8$, cut-off radius $r_c = 10$ and core radius $\gamma = 1$.

The mass ratios $\eta$ of the progenitor disks were varied between $\eta = 1$ and $\eta = 4$. For equal-mass mergers ($\eta = 1$) in total 400000 particles were adopted with each galaxy consisting of 20000 bulge particles, 60000 disk particles, and 120000 halo particles. Twice as many halo particles than disk particles are necessary in order to reduce heating and instability effects in the disk components [47]. For the mergers with $\eta = 2, 3, 4$ the parameters for the more massive galaxy were as described above. The low-mass companion however contained a fraction of $1/\eta$ less mass and number of particles in each component, with a disk scale length of $h = \sqrt{1/\eta}$, as expected from the Tully-Fisher relation [54].

The N-body simulations for the equal-mass mergers were performed by direct summation of the forces using the special purpose hardware GRAPE6 [44]. The mergers with mass ratios $\eta = 2, 3, 4$ were followed using the newly developed treecode WINE [64] in combination with the GRAPE5 [37] hardware. WINE uses a binary tree in combination with the refined multipole acceptance criterion proposed by Warren & Salmon (1996). This criterion enables the user to control the absolute force error which is introduced by the tree construction. We chose a value of 0.001 which guarantees that the error resulting from the tree is of order the intrinsic force error of the GRAPE5 hardware which is 0.1%. For all simulations we used a gravitational softening of $\epsilon = 0.05$ and a fixed leap-frog integration time step of $\Delta t = 0.04$. For the equal-mass mergers simulated with direct summation on GRAPE6 the total energy is conserved. The treecode in combination with GRAPE5 conserves the total energy up to 0.5%.

For all mergers, the galaxies approached each other on parabolic orbits with an initial separation of $r_{sep} = 30$ length units and a pericenter distance of $r_p = 2$ length units. Free parameters are the inclinations of the two disks relative to the orbital plane and the arguments of pericenter. In selecting unbiased initial parameters for the disk inclinations we followed the procedure described by Barnes [2]. To determine the spin vector of each disk we define four different orientations pointing to every vertex of a regular tetrahedron. These parameters result in 16 initial configurations for equal mass mergers and 16 more for each mass ratio $\eta = 2, 3, 4$ if the initial orientations are interchanged. In total we simulated 112 mergers.

In all simulations the merger remnants were allowed to settle into equilibrium approximately 8 to 10 dynamical times after the merger was complete. Then their equilibrium state was analysed.

# 3 Photometric and Kinematical Properties of the Remnants

To compare our simulated merger remnants with observations we analysed the remnants with respect to observed global photometric and kinematical properties of giant elliptical galaxies, e.g. surface density profiles, isophotal deviation from perfect ellipses, velocity dispersion, and major- and minor-axis rotation. Defining characteristic values for each projected remnant we followed as closely as possible the analysis described by Bender et al. [8].

## 3.1 Isophotal Shape

An artificial image of the remnant was created by binning the central 10 length units into $128 \times 128$ pixels. This picture was smoothed with a Gaussian filter of standard deviation 1.5 pixels. The isophotes and their deviations from perfect ellipses were then determined using a data reduction package kindly provided by Ralf Bender. Following the definition of Bender et al. [8] for the global properties of observed giant elliptical galaxies, we determined for every projection the effective $a_4$-coefficient $a4_{\mathbf{eff}}$ as the mean value of $a_4$ between $0.25r_e$ and $1.0r_e$, with $r_e$ being the projected spherical half-light radius. Like for observed ellipticals we find two types of remnants. Disky systems show a positive characteristic peak of $a_4$ roughly at $0.5r_e$. In boxy ellipticals, the $a_4$ coefficient might be positive in the innermost regions. It decreases however systematically outwards with a mean value that is negative. The characteristic ellipticity $\epsilon_{\mathbf{eff}}$ for each projection was defined as the isophotal ellipticity at $1.5r_e$. To investigate projection effects we determined for each simulation $a4_{\mathbf{eff}}$ and $\epsilon_{\mathbf{eff}}$ for 500 random projections of the remnant. These values were used to calculate the two-dimensional probability density function for a given simulated remnant to be "observed" in the $a4_{\mathbf{eff}}$-$\epsilon_{\mathbf{eff}}$ plane.

Figure 1 shows the ellipticities and $a_4$-coefficients of mergers with $\eta = 1, 2, 3,$ and 4. The contours indicate the areas of 50% (dashed line), 70% (thin line) and 90% (thick line) probability to detect a merger remnant with the given properties. Observed data points from Bender et al. [11] or [13] are over-plotted. Filled boxes are observed boxy ellipticals with $a4_{\mathbf{eff}} \leq 0$ while open diamonds indicate observed disky ellipticals with $a4_{\mathbf{eff}} > 0$. The error bar in determining $a_4$ from the simulations is shown in the upper left corner and was estimated applying the statistical bootstrapping method [36]. Ellipticity errors are in general too small to be visible.

We find that the isophotal shapes of ellipticals and their ellipticities are affected by the initial mass ratio of the merger and by projection effects. The area covered by 1:1 remnants with negative $a4_{\mathbf{eff}}$ is in very good agreement with the observed data for boxy elliptical galaxies. In particular the observed trend for more boxy galaxies to have higher ellipticities is reproduced. However we also find configurations of 1:1 mergers which under certain projection angles appear disky with $0 \geq a4_{\mathbf{eff}} \leq 1$. In addition, note that the remnants with $a4_{\mathbf{eff}}$ around zero can have higher ellipticities than observed.

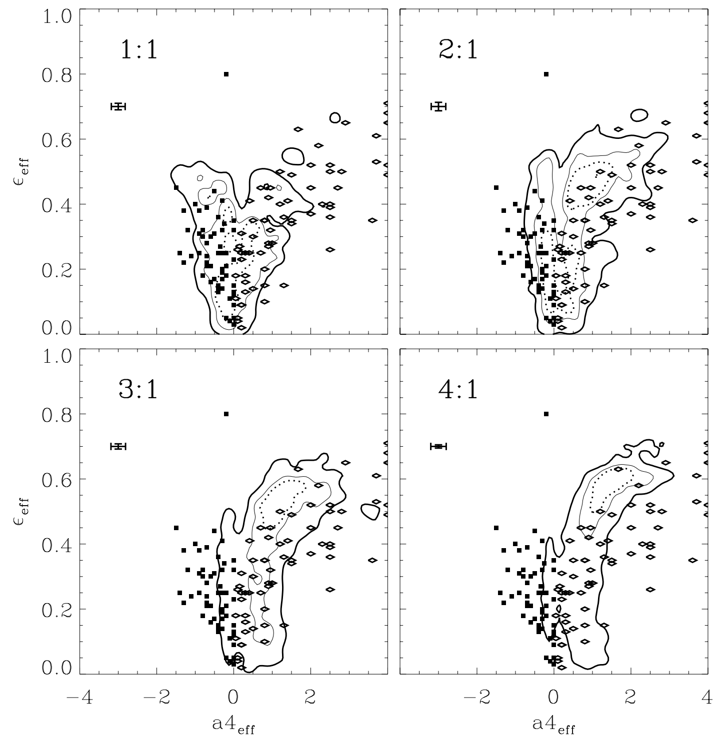



**Fig. 1.** Ellipticities versus fourth-order Fourier coefficient of the isophotal shape deviations is shown for simulations with different initial mass ratios. The contours indicate the 50% (dotted line), 70% (thin solid line) and the 90% (thick solid line) probability to find a merger remnant in the enclosed area. Black squares indicate values for observed boxy ellipticals, open diamonds show observed disky ellipticals.

The distribution function of isophotal shapes for 1:1 merger remnants peaks at $a4_{\mathbf{eff}} \approx -0.5\%$ (dashed curve in Fig. 2). It declines rapidly for more negative values and has a broad wing towards positive $a4_{\mathbf{eff}}$ values. Almost half of the projected remnants are disky. In contrast, remnants of mergers with higher mass ratios shift in the direction of positive $a4_{\mathbf{eff}}$. 2:1 remnants peak at $a4_{\mathbf{eff}} \approx 0$. Now, 75% of the projected remnants show disky isophotes. For these cases, the observed trend of more disky ellipticals to be more flattened is also clearly visible in Fig. 1. 3:1 and 4:1 mergers peak at $a4_{\mathbf{eff}} \approx 1$. Their fraction of boxy projections is only 11% and 7%, respectively. The very high positive values of $a4_{\mathbf{eff}} \geq 4\%$ observed in some ellipticals cannot be reproduced. One might argue that these objects formed from mergers with even higher mass ratios of $\eta \geq 5:1$. However, in this case, test simulations show that the merger remnants do not look like typical ellipticals anymore with characteristic de Vaucouleurs profiles as the more massive disk is not destroyed. Their surface brightness profiles instead remain exponential.

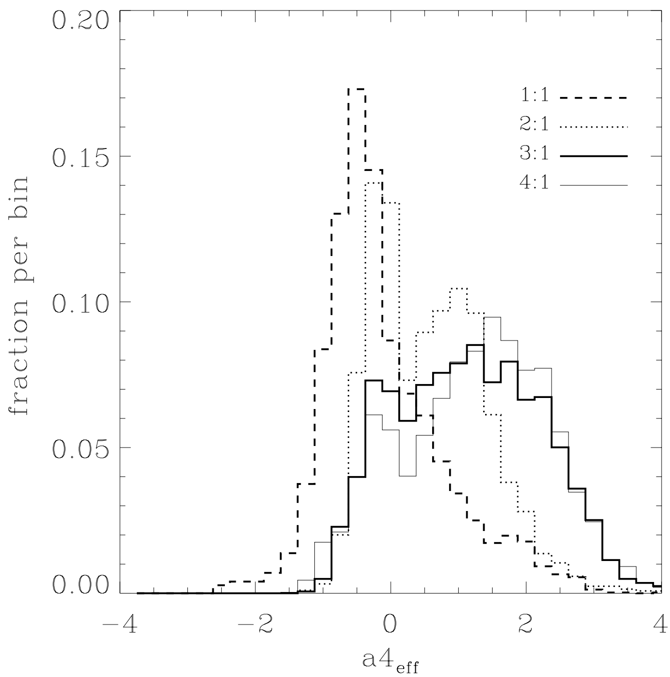



**Fig. 2.** Normalized histograms of the shape parameter $a4_{\mathbf{eff}}$ for mergers with various mass ratios.

In summary, there is a clear trend for unequal-mass mergers to produce more disky remnants. Responsible for the disky appearance of the 3:1 and 4:1 remnants is the distribution of the particles of the massive disk [3]. The particles originating from the small progenitor accumulate in a torus-like structure with peanut-shaped or boxy isophotes while the luminous material of the larger progenitor still keeps its disk-like appearance. In combination, the contribution from the larger progenitor – since it is three to four times more massive – dominates the overall appearance of the remnant. This result holds for all 3:1 and 4:1 merger remnants. For equal mass mergers however both disks are destroyed efficiently during the merger. No dominant disk-like structure remains after the merger and the system looses the information about the initial configuration.

### 3.2 Kinematics

The central velocity dispersion $\sigma_0$ of every remnant is determined as the average projected velocity dispersion of the stars inside a projected galactocentric radius of $0.2r_e$. The characteristic rotational velocity $v_{maj}$ along the major axis is defined as the projected rotational velocity determined around $1.5r_e$. Like for the isophotal shape we constructed probability density plots for the kinematical properties of the simulated remnants and compared them with observational

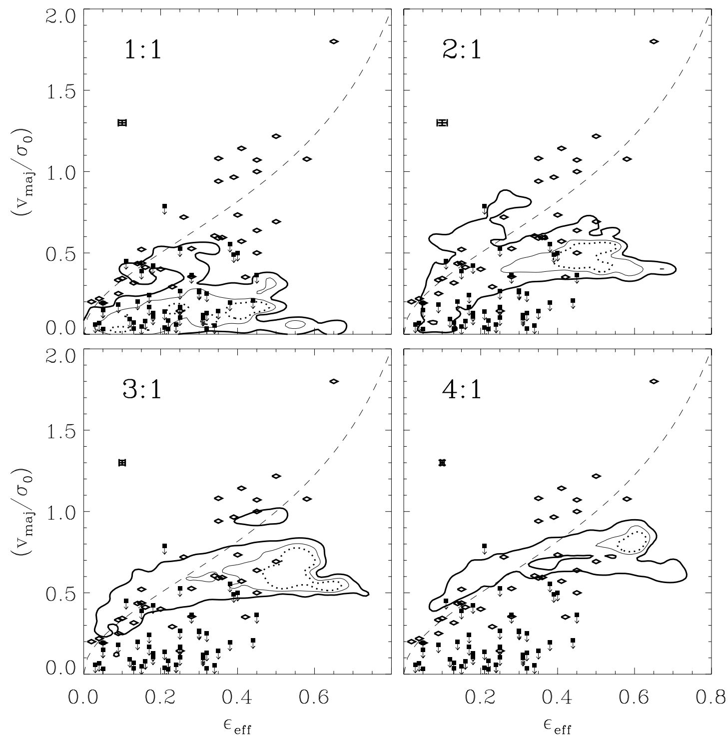



**Fig. 3.** Rotational velocity over velocity dispersion versus characteristic ellipticity for mergers with various mass ratios. Values for observed ellipticals are overplotted. The dashed line shows the theoretically predicted correlation for an oblate isotropic rotator.

data from elliptical galaxies. Figure 3 shows the distribution function in the $(v_{maj}/\sigma_0)$-$\epsilon_{\mathbf{eff}}$ plane.

The region of slowly rotating boxy ellipticals (filled squares) is almost completely covered by the data of 1:1 mergers. Unequal-mass merger remnants are clearly fast rotating. They can be associated with disky ellipticals. Although the simulated remnants are in good agreement with observations there is again the trend for the ellipticities to be higher than observed, especially when the system is seen edge-on.

The anisotropy parameter $(v_{maj}/\sigma_0)^*$ is defined as the ratio of the observed value of $(v_{maj}/\sigma_0)$ and the theoretical value for an isotropic oblate rotator $(v/\sigma)_{theo} = \sqrt{\epsilon_{\mathbf{obs}}/(1-\epsilon_{\mathbf{obs}})}$ with the observed ellipticity $\epsilon_{\mathbf{obs}}$ [17]. This parameter is frequently used by observers to test whether a given galaxy is flattened by rotation ($(v_{maj}/\sigma_0)^* \geq 0.7$) or by velocity anisotropy ($(v_{maj}/\sigma_0)^* < 0.7$) ([25], [7], [53], [58]). Figure 4 shows the normalized histograms for the $(v_{maj}/\sigma_0)^*$ values of the simulated remnants. 1:1 remnants peak around $(v_{maj}/\sigma_0)^* \approx 0.3$ with a more prominent tail towards lower values. They are consistent with being supported by anisotropic velocity dispersions. As these systems also have

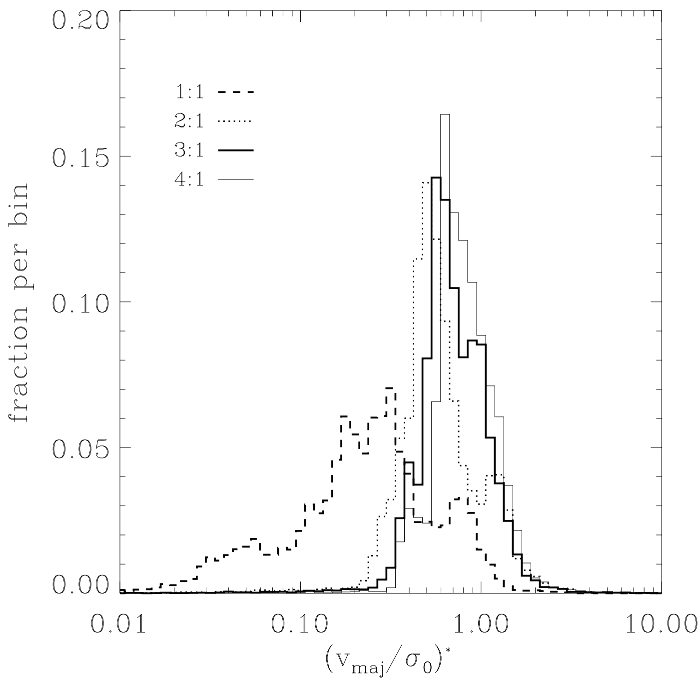



**Fig. 4.** Normalized histograms of $(v_{maj}/\sigma_0)^*$ for 1:1 (dashed line), 2:1 (dotted line), 3:1 (thick line) and 4:1 (thin line) mergers.

preferentially negative $a_4$-values they agree with observations of boxy ellipticals (Fig. 5). Unequal mass mergers peak at $(v_{maj}/\sigma_0)^* \approx 0.7$, as expected for oblate isotropic rotators. Since especially the 3:1 and 4:1 remnants also have predominantly disky isophotes they cover the area populated by observed disky ellipticals in the $log(v_{maj}/\sigma_0)^*$ - $a4_{\mathbf{eff}}$ diagram which is shown in Fig. 5.

We also investigated the minor-axis kinematics of the simulated remnants by determining the rotation velocity along the minor axis at $0.5r_{eff}$. The amount of minor axis rotation was characterized by $(v_{min}/\sqrt{v_{maj}^2 + v_{min}^2})$ [18]. Minor axis rotation in elliptical galaxies, in addition to isophotal twist, has been suggested as a sign for a triaxial shape of the main body of elliptical galaxies ([62], [30]). In general, 1:1 mergers show a significant amount of minor-axis rotation, whereas 3:1 and 4:1 remnants have only small minor axis rotation (for details see [50]).

## 4 Conclusions

The analysis of a large set of mergers with different mass ratios and orbital geometries shows that their properties are in general in good agreement with the observational data for elliptical galaxies.

Only equal mass mergers can produce boxy, anisotropic and slowly rotating remnants with a large amount of minor axis rotation. However, in the more un-

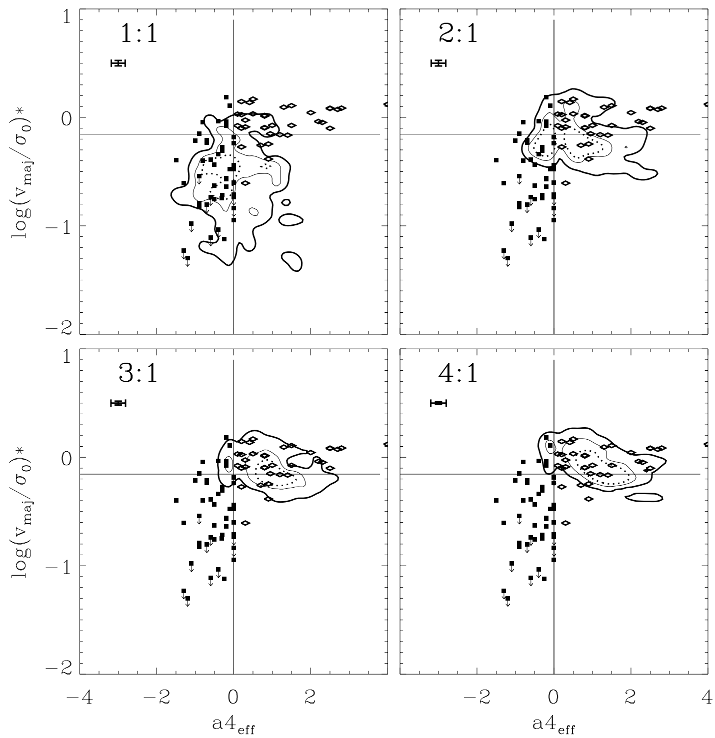



**Fig. 5.** Anisotropy parameter versus isophotal shape for mergers with various mass ratios. Values for observed ellipticals are overplotted.

likely case that the initial spins of the progenitor disks are aligned, the remnants appear isotropic and disky or boxy depending on the orientation. In contrast, 3:1 and 4:1 mergers form a more homogeneous group of remnants. They have preferentially disky isophotes, are always fast rotating and show small minor axis rotation independent of the assumed projection. 2:1 mergers have properties intermediate between boxy or disky ellipticals, depending on the projection and the orbital geometry of the merger.

There still exist problems which are not solved up to now. Certain projections of 1:1 mergers lead to anisotropic, disky remnants which are not observed. Edge-on projections of merger remnants often show very high ellipticities $\epsilon > 0.6$ which are larger than observed. Finally, some 2:1 to 4:1 remnants are more anisotropic than expected from their rotation. Their values of $(v/\sigma)^*$ are smaller and their ellipticities are larger than observed. A problem arises especially for very low luminosity giant ellipticals which are characterized by exceptionally high rotational velocities in the outer regions that cannot be reproduced [24]. A detailed analyses of the intrinsic kinematics of disky, fast rotating merger unequal-mass remnants which are called isotropic due to their high $(v/\sigma)^* \approx 0.7$ demonstrates that in most cases the velocity dispersion tensor is as anisotropic as for equal-mass, boxy and anisotropic mergers with $(v/\sigma)^* = 0.1$ [20]. The anisotropy parameter therefore is not necessarily a good indicator of anisotropy. It rather measures the amount of rotation in the systems.

The present simulations were purely dissipationless, taking into account only the stellar and dark matter components. The importance of gas in determining the structure of merger remnants is not clear up to now. Kormendy & Bender [41] proposed a revised Hubble sequence with disky ellipticals representing the missing link between late type systems and boxy ellipticals. They noted that gas infall into the equatorial plane with subsequent star formation could explain the origin of diskyness. Scorza & Bender [58] demonstrated that ellipticals with embedded disks would indeed appear disky when seen edge-on and boxy otherwise. Although this scenario appears attractive, it cannot explain why X-ray halos are found only in boxy ellipticals. As the detection of hot gas around galaxies should be independent of their orientation, the isophotal shapes of ellipticals would not correlate with their X-ray emission, if these shapes are merely a result of projection effects.

Our simulations indicate that it is preferentially the initial mass ratio which determines the isophotal shapes of merger remnants. Still, gas could have played an important role in affecting the final structure and stellar population of ellipticals ([5], [6], [46]), not only in their central regions and might solve the problem of dissipationless mergers. Naab & Burkert [48] have shown that extended gas disks can form as a result of a gas rich unequal mass mergers (see also [4]). Naab & Burkert [49] investigated line-of-sight velocity distributions of dissipationless merger remnants and found a velocity profile asymmetry that is opposite to the observations. They concluded that this disagreement can be solved if ellipticals would indeed contain a second disk-like substructure that most likely formed through gas accretion. The situation is however not completely clear, as another study by Bendo & Barnes [15] found a good agreement of the observed asymmetries for some cases. More simulations, including gas and star formation will be required to understand the role of gas in mergers and to answer the question of how early-type galaxies formed.

**Acknowledgement**

A. Burkert thanks the organizers of the workshop "Galaxies and Chaos" for the invitation and a very stimulating conference.

## References

1. Aragon-Salamanca, A. , Ellis, R. S., Couch, W. J. & Carter, D. 1993, MNRAS, 262, 764
2. Barnes, J. E. 1988, ApJ, 331, 699
3. Barnes, J. E. 1998, Galaxies: Interactions and Induced Star Formation: Lecture Notes 1996 / Saas Fee Advanced Course 26, eds. D. Friedli, L. Martinet, and D. Pfenniger, Springer, 275
4. Barnes, J. E. 2002, MNRAS, 333, 481

5. Bekki, K. 1998, ApJL, 502, L133
6. Bekki, K. 1999, ApJ, 513, 108
7. Bender, R. 1988a, A&A, 193, L7
8. Bender, R. 1988b, A&A, 202, L5
9. Bender, R., Döbereiner, S., & Möllenhoff, C. 1988, A&AS, 74, 385
10. Bender, R., Surma, P., Döbereiner, S., Möllenhoff, C. & Madejsky, R. 1989, A&A, 217, 35
11. Bender, R. & Surma, P. 1992, A&A, 258, 250
12. Bender, R. 1996, IAU Symp., 171, New Light on Galaxy Evolution, ed. R. Bender & R. L. Davies (Dordrecht: Kluwer), 181
13. Bender, R., Burstein, D. & Faber, S.M. 1992, ApJ, 399, 462
14. Bender, R., Ziegler, B. & Bruzual, G. 1996, ApJL, 463, L51
15. Bendo, G. J. & Barnes, J. E. 2000, MNRAS, 316, 315
16. Beuing, J., Döbereiner, S., Böhringer, H. & Bender, R. 1999, MNRAS, 302, 209
17. Binney, J. 1978, MNRAS, 183, 779
18. Binney, J. 1985, MNRAS, 212, 767
19. Burkert, A. 1993, A&A, 278, 23
20. Burkert, A., Binney, J. & Naab, T. 2003,in preparation
21. Bower, R. G., Lucey, J. R. & Ellis, R. S. 1992, MNRAS, 254, 589
22. Bruzual A., G. & Charlot, S. 1993, ApJ, 405, 538
23. Carlberg, R. G. 1986, ApJ, 310, 593
24. Cretton, N., Naab, T., Rix, H., & Burkert, A. 2001, ApJ, 554, 291
25. Davies, R. L., Efstathiou, G., Fall, S. M., Illingworth, G. & Schechter, P. L. 1983, ApJ, 266, 41
26. Davies, R. L. 1996, IAU Symp., 171, New Light on Galaxy Evolution, ed. R.Bender & R. L. Davies (Dordrecht: Kluwer), 37
27. Ellis, R. S., Smail, I., Dressler, A. , Couch, W. J., Oemler, A. , Jr., Butcher, H. & Sharples, R. M. 1997, ApJ, 483, 582
28. Faber, S. M., et al. 1997, AJ, 114, 1771
29. Franx, M. & Illingworth, G. D. 1988, ApJL, 327, L55
30. Franx, M., Illingworth, G. & de Zeeuw, T. 1991, ApJ, 383, 112
31. Gerhard, O.E. 1981, MNRAS 197, 179
32. Hernquist, L. 1990, ApJ, 356, 359
33. Hernquist, L. 1992, ApJ, 400, 460
34. Hernquist, L. 1993a, ApJS, 86, 389
35. Hernquist, L. 1993b, ApJ, 409, 548
36. Heyl, J. S., Hernquist, L. & Spergel, D. N. 1994, ApJ, 427, 165
37. Kawai, A., Fukushige, T., Makino, J., & Taiji, M. 2000, PASJ, 52, 659
38. Kauffmann, G. 1996, MNRAS, 281, 487
39. Kauffmann, G. & Charlot, S. 1998, MNRAS, 297, L23
40. Khochfar, S., Burkert, A. & White. S. 2003, in preparation
41. Kormendy, J. & Bender, R. 1996, ApJL, 464, L119
42. Larson, R.B. 1974, MNRAS, 166, 585
43. Lauer, T. R., et al. 1995, AJ, 110, 2622
44. Makino, J., Fukushige, T. & Namura, K. 2003, to be submitted to PASJ.
45. Mehlert, D., Saglia, R. P., Bender, R. & Wegner, G. 1998, A&A, 332, 33
46. Mihos, J. C. & Hernquist, L. 1996, ApJ, 464, 641
47. Naab, T. , Burkert, A., & Hernquist, L. 1999, ApJL, 523, L133
48. Naab, T. & Burkert, A. 2001a, ASP Conf. Ser. 230: Galaxy Disks and Disk Galaxies, 451

49. Naab, T. & Burkert, A. 2001b, ApJL, 555, L91
50. Naab, T. & Burkert A., submitted to ApJ
51. Navarro, J.F. & Steinmetz, M. 2000, ApJ, 538, 477
52. Negroponte, J. & White, S. D. M. 1983, MNRAS, 205, 1009
53. Nieto, J.-L., Capaccioli, M. & Held, E. V. 1988, A&A, 195, L1
54. Pierce, M. J. & Tully, R. B. 1992, ApJ, 387, 47
55. Rix, H. -W. & White, S. D. M. 1990, ApJ, 362, 52
56. Rix, H. -W. & White, S. D. M. 1992, MNRAS, 254, 389
57. Jedrzejewski, R. & Schechter, P. L. 1988, ApJL, 330, L87
58. Scorza, C. & Bender, R. 1995, A&A, 293, 20
59. Searle, L., Sargent, W. L. W. & Bagnuolo, W. G. 1973, ApJ, 179, 427
60. Steinmetz, M. & Buchner, S. 1995, Galaxies in the Young Universe, Proceedings of a Workshop held at Ringberg Castle, eds. H. Hippelein, K. Meisenheimer & H.-J. Röser, Springer, p. 215
61. Toomre, A. & Toomre, J. 1972, ApJ, 178, 623
62. Wagner, S. J., Bender, R. & Moellenhoff, C. 1988, A&A, 195, L5
63. Warren, M. S., & Salmon, J. K. 1996, ApJ, 460, 121
64. Wetzstein, M., Nelson, A., Naab, T., & Burkert, A. 2003, in preparation
65. Ziegler, B. L. & Bender, R. 1997, MNRAS, 291, 527
66. Ziegler, B. L., Saglia, R. P., Bender, R., Belloni, P., Greggio, L. & Seitz, S. 1999, A&A, 346, 13

# Dynamical Evolution of Galaxies: Supercomputer $N$-Body Simulations

Edward Liverts, Evgeny Griv, Michael Gedalin, and David Eichler

Ben-Gurion University of the Negev, P.O. Box 653, Beer-Sheva 84105, Israel

**Abstract.** The time evolution of a computer model for an isolated disk representing a flat galaxy is studied. The method of direct integration of Newton's equations of motion of particles–"stars" is applied. Using the modern 128-processor SGI Origin 2000 supercomputer in Israel, we make long simulation runs with a large number of particles, $N = 100\,000$. One of the goals of the simulation is to test the validities of the modified Safronov–Toomre criterion for stability of arbitrary but not only axisymmetric Jeans-type gravity disturbances (e.g., those produced by a spontaneous perturbation and/or a companion system) in a self-gravitating, thin, and almost collisionless stellar disk. We are also interested in how model particles diffuse in chaotic (residual) velocity space. This is of considerable interest in the nonlinear theory of stellar disks.

## 1 Introduction

One can learn much about the properties of stellar systems of galaxies experimentally by computer simulation of $N$-body systems. In this work, we analyze the evolution and stability of structures in an $N$-body model of an isolated and rotating stellar disk representing a flat galaxy by integration of Newton's equations of motion of $N$ identical particles. Use of the 128-processor SGI Origin 2000 computer, enabled us to make long simulation runs using a large number of particles, $N = 100\,000$, in the direct summation code and thus simulate phenomena not previously studied numerically. The essential difference between the present and previous simulations is the comparison between the results of $N$-body experiments and the kinetic stability theory as developed in [1–8].

## 2 $N$-Body Simulations

Different methods are currently employed to simulate the evolution of collisionless point-mass systems of flat galaxies by $N$-body experiments. See, e.g., [9] as a review. For instance, one can use an algorithm for a simulation code, which is an analog of plasma particle-mesh (PM) codes. It is believed that simulating many billions of stars in actual galaxies by using only several ten or hundred thousands particles in PM experiments will be enough to capture the essential physics, which includes wave-like collective motions. In other fields, such as the simulation of spiral structures, PM codes may be used with moderate success. This is because these fine-scale $\lesssim 1$ kpc structures can basically be governed by collisionless processes. By increasing the number of cells to reproduce the

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 340–347, 2003.

microstructures, one reduces the average number of particles per cell, and thus increases the undesirable effect of particle encounters. The problem is serious because there are regions in the phase space which are important to the problem but in which the distribution function is small. In contrast, our model is based on the direct numerical integration of the equations of motion in three dimensions for $N$ mutually gravitating particles.

The numerical procedure is used first to seek stationary solutions to the Boltzmann equation in the self-consistent field approximation, and then to determine the stability of those solutions to small gravity perturbations. At the start of the $N$-body integration, our similation initilizes the particles on a set of 100 concentric circular rings with a circular velocity $\boldsymbol{V}_{\mathrm{rot}}$ of galactic rotation in the equatorial plane; the system is isolated. Then the position of each particle was slightly perturbed by applying a pseudorandom number generator. The Maxwellian-distributed chaotic (residual) velocities $\boldsymbol{v}$ were added to the initial circular velocities $\boldsymbol{V}_{\mathrm{rot}}$, and $|\boldsymbol{v}| \ll |\boldsymbol{V}_{\mathrm{rot}}|$. The acceleration of the $i$th particle is

$$\boldsymbol{a}_i = \sum_{j \neq i}^{N} \frac{(\boldsymbol{r}_j - \boldsymbol{r}_i)}{(r_{ij}^2 + r_{\mathrm{cut}}^2)^{3/2}} . \tag{1}$$

In (1), $r_{\mathrm{cut}}$ is the so-called cutoff radius. This "softening" of the gravitational potential is a device used in $N$-body simulations to avoid numerical difficulties at very close but rare encounters. Units are chosen such that the mass of each particle is $10^5 M_\odot$ so that the total mass of the disk galaxy is $10^{10} M_\odot$. The initial radius of the disk is 10 kpc. When $N$ is large, the main computational problem is the large number, $\propto N(N-1)$, of operations required to determine $\boldsymbol{a}_i$.

We consider a rotating model disk of stars of thickness $h$ with a surface mass density variation given by $\sigma_0(r) = \sigma(0)\sqrt{1 - r^2/R^2}$, where $\sigma(0)$ is the central surface density and $R$ is the radius of the initial disk. As a solution of a time-independent collisionless Boltzmann equation, to ensure initial equilibrium, the angular velocity to balance the zero-velocity dispersion disk, $\Omega_0 = \pi\sqrt{G\sigma(0)/2R}$, was adopted [10]. For this uniformly rotating disk, the Maxwellian-distributed chaotic velocities with radial $c_r$ and azimuthal $c_\varphi$ dispersions in the plane $z = 0$, according to the Safronov–Toomre criterion [11], $c_{\mathrm{T}} = 3.4G\sigma_0/\kappa = 0.341\Omega_0\sqrt{R^2 - r^2}$, may be added to the initial circular velocities $V_{\mathrm{rot}} = r\Omega_0$. Here $\kappa$ is the epicyclic frequency. According to [1–8, 12–14], it is crucial to realize that such a spatially inhomogeneous disk is Jeans-stable only against the axisymmetric (radial) gravity perturbations but unstable against the nonaxisymmetric (spiral) perturbations. The initial vertical velocity dispersion was chosen $c_z = 0.15c_r$. Finally, the angular velocity $\Omega_0$ was replaced by [10]

$$\Omega = \left\{ \Omega_0^2 + \frac{1}{r\sigma(r)} \frac{\partial}{\partial r} \left[ \sigma(r) c_r^2(r) \right] \right\}^{1/2} .$$

The sense of disk rotation was taken to be counterclockwise, the cutoff radius was $r_{\mathrm{cut}} = 0.004R$, and the initial disk thickness was $h = 0.006R$, that is, $h > r_{\mathrm{cut}}$.

Slight corrections have been applied to the resultant velocities and coordinates of the model stars so as to ensure the equilibrium between the centrifugal and gravitational forces, to preserve the position of the disk center of gravity at the origin, and to include the weak effect of the finite thickness of the disk to the gravitational potential. Thus, the initial model is very near the dynamical equilibrium for all radii. A time $t = 1$ was taken to correspond to a single revolution of the initial disk. In the experiment the simulation was performed up to a time $t = 10$. It should be noted here that after about three rotations the picture is practically stabilized and no significant changes in gross properties of the model over this time are observed. Tests indicated that the results were insensitive to changes in the number of particles in the range $N = 10\,000 - 100\,000$, the cutoff parameter in the range $r_{\mathrm{cut}} = (0.001 - 0.01)R$, the initial velocity dispersion in the range $c_r = (1 - 1.3)c_{\mathrm{T}}$, and the initial disk thickness in the range $h = (0 - 0.08)R$. We argue that structures observed in our $N$-body simulations originate from the collective modes of oscillations — the classical Jeans-type gravitational modes and firehose-type bending modes.

## 3 Results of Simulations

Figure 1 displays a series of snapshots from a simulation run. The figures include only a face-on view of the simulation region. In accordance with the theoretical explanation [1–8, 12–14], the effects of the Jeans instability of spontaneous non-axisymmetric gravity perturbations appear quickly in the simulation. One can see at first a strong multi-armed spiral structure. It is interesting to notice that in a sample of 654 optical spiral galaxies [15], two-armed (grand design) galaxies like M 51 are roughly a factor of six times rare than such many-armed galaxies like NGC 613, an SBb galaxy in Sculptor.

At a later time, $t > 0.5$, the multi-armed structure disappears quickly and is replaced by a weak spiral structure with three main spiral arms, $m = 3$, or sometimes two, $m = 2$, or only one, $m = 1$, spiral arms. These spirals are evidently gravitationally (Jeans-)unstable Lin–Shu density waves [16–19] and not material arms, since test particles pass right through them. The $m = 1$ mode shifts the point with highest density from the center of mass [9]. Interestingly, in many disk-shaped galaxies, e.g., in the spiral galaxies M 101 and NGC 1300,

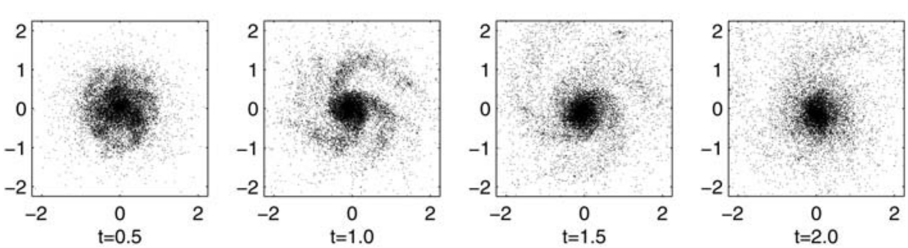



**Fig. 1.** The time evolution (face-on view) of an initially equilibrium, Toomre-stable disk of $N = 100\,000$ stars. The effects of the Jeans-type gravitational instability of spontaneous spiral disturbances appear quickly in the simulation. A moderately tightly wound, low–$m$ spiral structure develops in the plane of the system at a time $t \approx 1.0$.

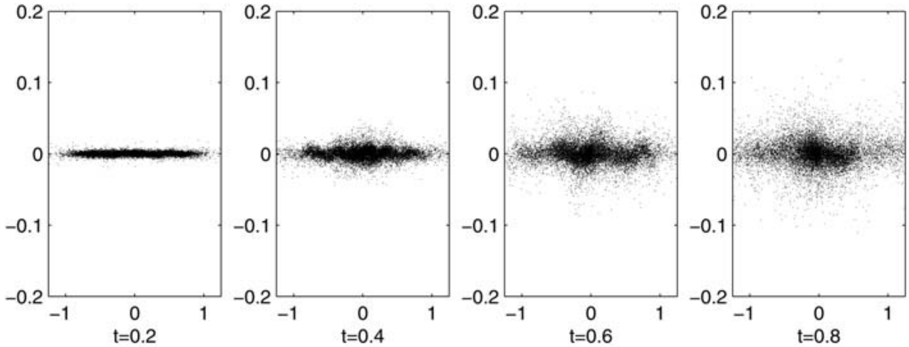



**Fig. 2.** The time evolution (edge-on view) for the simulation run shown in Fig. 1. At $t \approx 0.4$, the firehose-type bending instability fiercely develops in the central, almost nonrotating parts of the system (and is switched off at $t \approx 2.0$ [9, 21]).

there appears to be a deviation from rotational symmetry. In principle, such a deviation may be due to the one arm Jeans instability. Also note that two unusual single-arm galaxies turn up in [20] sample of some 54 differentially rotating spiral galaxies. In one of them, namely NGC 4378, the spiral arm can be traced over most $\frac{1}{4}$ revolutions.

From the edge-on view pictured in Fig. 2, one can see that a fully three-dimensional disk develops immediately at $t \approx 0.2$. A straightforward estimate shows that a mean height $\Delta$ of the disk above the plane corresponds to the force balance between the gravitational attraction in the plane and the "pressure" due to the velocity dispersion $c_z$ (i.e., "temperature") in the $z$-direction [9, 19]. Clearly, this pressure-supported (in the $z$-direction) three-dimensional structure seen to form very rapidly on the time scale of a single vertical epicyclic oscillation, $< \Omega^{-1}$, with rather sharp edges. After a time $t \approx 0.2$ there is no change in the edge-on structure until at $t \approx 0.4$. It is noteworthy that at $t \approx 0.4$, the firehose-type bending instability rapidly develops in the central parts of the system (and is switched off at $t \approx 2.0$ [9, 21]). At later times, $t > 2.0$, no dramatic evolution is observed in our simulations (see [9]). To emphasize, the bending instability develops in the *central*, almost *nonrotating* region of the system under study [9]. New spectroscopic optical and H I observations constitute a strong case in favour of this bar-buckling mechanism for the formation of boxy/peanut-shaped bulges in spiral galaxies [22]. Apparently, the authors of [23–25] first found the firehose-type bending instability as a precursor of galactic bulge formation in the central, almost nonrotating regions of a warm in the plane $N$-body disk, which initially developed planar bars.

At somewhat later times, $t > 2.0$, a "box-shaped" or sometimes "peanut-shaped" bar structure is developed [9]. The simulations show that, soon after a central bar develops in the equatorial plane, it buckles and settles with an increased thickness and vertical velocity dispersion, appearing boxy-shaped when seen end-on and peanut-shaped when seen side-on [9, 24, 25]. The projected

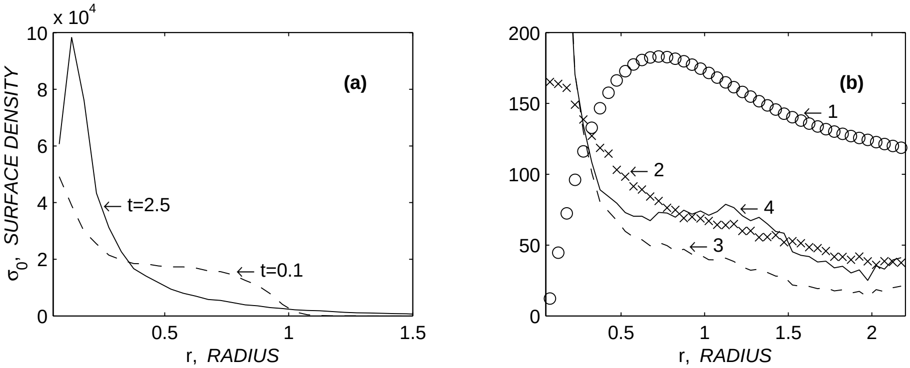



**Fig. 3.** In Fig. 3*a*, we show the evolution of the surface density for a stellar disk ($\times 2M_{\odot}/\mathrm{kpc}^2$) shown in Fig. 1. In Fig. 3*b*, rotation curve $V_{\mathrm{rot}}$ (km/sec) (*1*), radial velocity dispersion (km/sec) (*2*), Safronov–Toomre critical velocity dispersion (km/sec) (*3*), and modified dispersion (km/sec) (*4*) at the time $t = 2.5$.

densities of this central bar resemble the bulge light distribution measured by the COBE satellite in the Milky Way's Galaxy [26]. Moreover, a significant fraction of edge-on spiral galaxies, and therefore presumably of all spirals, show boxy or peanut-shaped isophotes in the bulge region [22]. The firehose-type bar-buckling instability is the currently favored mechanism for the formation of boxy/peanut-shaped bulges in spiral galaxies.

Figure 3*a* shows the evolution of the azimuthally averaged surface density $\sigma_0$ as a function of radius. It can be seen that the mass density is redistributed by the Jeans-unstable waves on the dynamical time scale, $\lesssim \Omega^{-1}$; the surface density of the quasi-steady state system at $t > 1$ falls off exponentially.

### 3.1 Modified Stability Criterion

To emphasize it again, even though the initial velocity dispersion is equal to the Safronov–Toomre [11] stabilizing one $c_{\mathrm{T}}$, the model is still violently Jeans-unstable. The reason for such a behaviour of a disk has been explained in [1–8]. See also [12–14, 27–30] for a discussion. Accordingly, the presence of the differential rotation (or shear) results in quite different dynamical properties of the axisymmetric and nonaxisymmetric gravity perturbations. In differentially rotating disks the azimuthal force resulting from azimuthal displacements is more important in determining the stability than is the radial force resulting from radial displacements [28, 31]. In a nonuniformly rotating disk for nonaxisymmetric perturbations the modified dispersion $c_{\mathrm{M}}$ of a marginally Jeans-stable system is larger than $c_{\mathrm{T}}$ (although still of the order of $c_{\mathrm{T}}$), and is approximately

$$c_{\mathrm{M}} \approx (2\Omega/\kappa) c_{\mathrm{T}} . \tag{2}$$

In disk-shaped galaxies, $2\Omega/\kappa = 1.5 - 1.8$. It is obvious that in differentially rotating galaxies, disks manage to keep their local stability parameter close to the critical value, $c_r \approx (2\Omega/\kappa)c_{\mathrm{T}} \approx 2c_{\mathrm{T}}$ or Toomre's $Q$-stability parameter $Q \equiv c_r/c_{\mathrm{T}} \approx 2\Omega/\kappa \approx 2$, respectively. In this case, once the entire differentially rotating disk has been heated to values $c_r \approx 2c_{\mathrm{T}}$ (or $Q \approx 2$), no further spiral waves can be sustained by virtue of the Jeans instability — unless some "cooling" mechanism is available leading to Toomre's $Q$-value under approximately 2 or to the value of $c_r$ smaller than approximately $2c_{\mathrm{T}}$, respectively (e.g., by the dissipation in the gas and/or by the star formation in an interstellar medium [9]).

Thus, the Jeans-unstable perturbations can be stabilized by the chaotic velocity spread. The critical Safronov–Toomre [11] velocity dispersion should stabilize only radial perturbations of the Jeans type. The differentially rotating and spatially inhomogeneous disk is still unstable against spiral Jeans perturbations. The modified stability criterion against arbitrary but not only axisymmetric gravity perturbations is given by (2). The spiral arms in nonuniformly rotating systems are a mechanism for angular momentum transfer [31].

We compare the radial velocity dispersion values $c_r$ predicted by the Safronov–Toomre criterion $c_{\mathrm{T}}$ and by the modified criterion (2) with values obtained in the numerical experiment. Many investigators have remarked that the experimental $c_r$ significantly exceeds $c_{\mathrm{T}}$. This effect is also apparent in Fig. 3*b*, which represents the quasi-steady state for the computer experiment with a disk shown in Fig. 1. On the other hand, the quantity $c_{\mathrm{M}}$ calculated from (2) satisfactorily fits the experimental $c_r$ (in the central, pressure-supported parts of the system where $V_{\mathrm{rot}} \lesssim c_r$, both $c_{\mathrm{T}}$ and $c_{\mathrm{M}}$ fail to employ). Thus, the experiment yields a radial velocity dispersion for the particles significantly greater than predicted by the Safronov–Toomre criterion, but it is nearly equal to the modified dispersion.

### 3.2 Chaotic Velocity Diffusion

One of the important problems of stellar disks is the determination of the chaotic velocity diffusion. Such a velocity diffusion can be caused by gravitational instabilities of a disk. To compute the velocity diffusion we calculate the mean-square spread in the planar chaotic velocity $c^2$ as a function of time for different radii.

As is seen in Fig. 4, along with the growth of the oscillation amplitude (planar spiral density waves): (a) chaotic velocities increase, and eventually in the disk a quasi-stationary distribution is established at times $t \gtrsim 1$ so that the Jeans stability sets in, and (b) during the first rotation the squared plane velocity dispersion of particles increases with time as roughly $c^2 \equiv c_r^2 + c_\varphi^2 \propto t$. The results (a) and (b) are in good agreement with the predictions of weakly nonlinear (quasilinear) kinetic theory as developed in [7, 8]. Interestingly, observations already showed about the same law of "heating" (increase of the velocity dispersion of the young stellar population with age of stars $t$) in the Solar vicinity, $c^2 \propto t$ [32]. We conclude that both the quasilinear theory [7, 8] and the $N$-body simulation presented here are able to account for the form of the age–velocity dispersion law in the plane of the Galaxy.

In turn, there are numerous observations showing that there exists ongoing dynamical relaxation on the time scale of $< 10$ rotation periods ($< 2 \times 10^9$ yr)

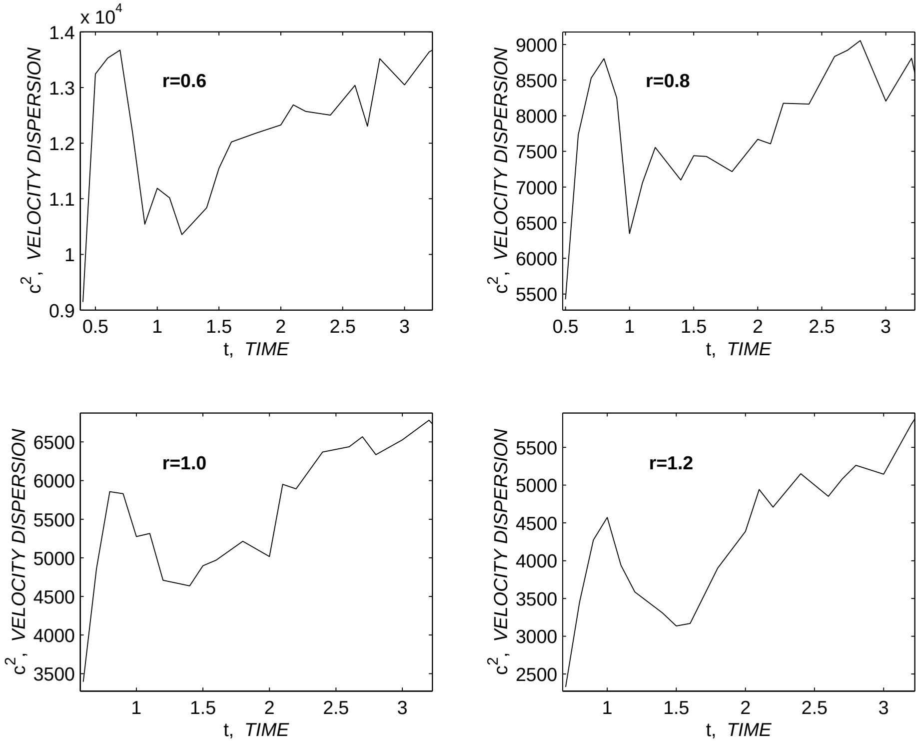



**Fig. 4.** The time evolution of the squared planar velocity dispersion $c^2$ for different radii $r$ ($\times 10$ kpc). Initially, the dispersion increases with with time rapidly, $c^2 \propto t$. Later in a quasi-steady state, $t \gtrsim 1$, the dispersion grows only slightly.

in the collisionless disk of the Milky Way's Galaxy [19, 32–34]. It was observed that in the Solar neighborhood the velocity distribution function of stars with an age $t \gtrsim 10^8$ yr is close to a Schwarzschild distribution — a set of Gaussian distributions along each coordinate in chaotic velocity space, i.e., close to equilibrium along each coordinate. Also, older stellar populations are observed to have a higher velocity dispersion than younger ones. Thus, this dynamical relaxation of the distribution of young stars which were born in the equilibrium disk of the Galaxy results in a randomization of the velocity distribution and a monotonic increase of the chaotic velocity dispersion. The latter indicates a significant irregular gravitational field in the Galactic disk [32–34]. The irregular field causes a diffusion of stellar orbits in velocity [7] (and positional [8]) space. Various mechanisms for the relaxation have been proposed. In the present work, we suggest the idea of the collective *collisionless* relaxation: planar Jeans-unstable gravity perturbations affect effectively the averaged velocity distribution of young stars in the equatorial plane (see [7, 8, 35, 36] for a discussion).

We thank Tzi-Hong Chiueh, Frank H. Shu, Irina Shuster, Raphael Steinitz, and Chi Yuan for valuable discussions. This work was supported in part by the Israel Science Foundation and the Israeli Ministry of Immigrant Absorption.

## References

1. E. Griv, W. Peter: ApJ **469**, 84 (1996)
2. E. Griv, M. Gedalin, C. Yuan: A&A **328**, 531 (1997)
3. E. Griv, B. Rosenstein, M. Gedalin, D. Eichler: A&A **347**, 821 (1999)
4. E. Griv, C. Yuan, M. Gedalin: MNRAS **307**, 1 (1999)
5. E. Griv, M. Gedalin, D. Eichler, C. Yuan: Phys. Rev. Lett. **84**, 4280 (2000)
6. E. Griv, M. Gedalin, D. Eichler, C. Yuan: Ap&SS **271**, 21 (2000)
7. E. Griv, M. Gedalin, D. Eichler: ApJ **555**, L29 (2001)
8. E. Griv, M. Gedalin, C. Yuan: A&A **383**, 338 (2002)
9. E. Griv, T. Chiueh: ApJ **503**, 186 (1998)
10. F. Hohl: J. Comput. Phys. **9**, 100 (1972)
11. A. Toomre: ApJ, **139**, 1217 (1964)
12. A.G. Morozov: SvA **24**, 391 (1980)
13. A.G. Morozov: SvA **25**, 421 (1981)
14. E. Liverts, E. Griv, D. Eichler, M. Gedalin: Ap&SS **274**, 315 (2000)
15. B.G. Elmegreen, D.M. Elmegreen: ApJ **342**, 677 (1989)
16. C.C. Lin, F.H. Shu: Proc. Natl. Acad. Sci. **55**, 229 (1966)
17. C.C. Lin, C. Yuan, F.H. Shu: ApJ **155**, 721 (1969)
18. F.H. Shu: ApJ **160**, 99 (1970)
19. J. Binney, S. Tremaine: *Galactic Dynamics* (Princeton Univ. Press, Princeton, NJ 1987)
20. J. Kormendy, C.A. Norman: ApJ **233**, 539 (1979)
21. E. Griv, M. Gedalin, C. Yuan: ApJ **580**, L27 (2002)
22. M. Bureau, K.C. Freeman: AJ **118**, 126 (1999)
23. F. Combes, R.H. Sanders: A&A **96**, 164 (1981)
24. F. Combes, F. Debbasch, D. Friedli, D. Pfenniger: A&A **233**, 82 (1991)
25. N. Raha, J.A. Sellwood, R.A. James, F.D. Kahn: Nature **352**, 411 (1991)
26. E. Dwek, R.G. Arendt, M.G. Hauser et al.: ApJ **445**, 716 (1995)
27. Y.Y. Lau, G. Bertin: ApJ **226**, 508 (1978)
28. C.C. Lin, Y.Y. Lau: SIAM Stud. Appl. Math. **60**, 97 (1979)
29. G. Bertin: Phys. Rep. **61**, 1 (1980)
30. C.C. Lin, G. Bertin: Adv. Appl. Mech. **24**, 155 (1984)
31. D. Lynden-Bell, A.J. Kalnajs: MNRAS **157**, 1 (1972)
32. R. Wielen: A&A **60**, 263 (1977)
33. G. Gilmore, I.R. King, P.C. van der Kruit: *The Milky Way as a Galaxy* (Univ. Science Book, Mill Valley, CA 1990)
34. J. Binney: 'Secular Evolution of the Galactic Disk'. In: *Galaxy Disks and Disk Galaxies, ASP Conf. Series, **230***, ed. J.G. Funes, E.M. Corsini (PASP, San Francisco 2001), pp. 63–70
35. E. Griv, M. Gedalin, D. Eichler, C. Yuan: 'The Dynamical Relaxation of the Milky Way'. In: *Astrophysical Ages and Times Scales, ASP Conf. Series, **245***, ed. T. von Hippel, C. Simpson, N. Manset (PASP, San Francisco 2001), pp. 280–287
36. E. Griv, M. Gedalin, C. Yuan: A&A (2003), in press

# Formation of the Halo Stellar Population in Spiral and Elliptical Galaxies

Tetyana Nykytyuk

Main Astronomical Observatory of Academy of Sciences of Ukraine, 03680 Kyiv, Zabolothnoho 27, Ukraine

**Abstract.** A scenario of galactic halo formation through mergers of fragments has been considered. In the framework of the scenario sets of fragments have been obtained from the observed halo metallicity distribution function of the Milky Way Galaxy and others. Our results allow us to conclude that 1) in our Galaxy a halo field star formation can be occured in the fragments evolved as closed system 2) the formation of the bulk of halo field stars of M31 and NGC 5128 perhaps is not associated with the formation of the halo globular clusters in these galaxies 3) in our Galaxy the formation of the halo field stars could be associated with the halo globular cluster formation

## 1 Introduction

The oldest stellar systems carry information about the processes that took place in galaxies during the early epoch of their formation. Studies of ages and metallicities of these systems allow us to put restrictions on the theoretical description of processes of galaxy formation. Therefore the halo stellar population is good test for models of protogalaxy formation.

Studying globular clusters of our Galaxy Searle and Zinn [18] have concluded that the halo globular clusters were formed in fragments which have merged with the main body on timescales of more than 1 Gyr. Our work is based on this scenario. It is supposed that the halo stellar population represents a mixture of stars that were formed in fragments originally evolved separately from the main protogalactic cloud. Hence, there should be a set of fragments which will reproduce the observed halo metallicity distribution of stars. The aim of this work is to identify a set of fragments containing stars which when mixed give rise to to the observed metallicity distribution of halo field stars and of halo globular clusters.

## 2 The Model

Let the mass of fragments from which the halo is formed equal the sum of the masses of the stellar population and of gas fallen onto the disk at the present epoch. Each fragment is supposed to evolve as a closed system; mergers among fragments are not taken into account. Star formation process in different fragments can begin at different times. A fragment may evolve up to given astration level s ($s = 1 - \mu$, $\mu$ - the fraction of gas in a fragment) before it falls on a protogalaxy. The stars formed up to this moment add to the halo stellar population

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 348–356, 2003.

and the gas falls onto the disk. Fragments can be captured by the protogalaxy untill they begin to form an internal stellar population, i.e. only gas fragments are captured. The contribution of elements, synthesized by a stellar population that formed in one fragment, to interstellar medium depends on 1) synthesis of elements by stars of various masses 2) the number of stars formed in a given interval of stellar masses i.e. on the initial mass function. The initial mass function in a mass range from m up to $m + dm$ is

$$N(m) = \phi_0 m^{-A} dm, \tag{1}$$

where $\phi_0$ is the normalization coefficient determined from the condition

$$\phi_0 \sum_{j=1}^{n} m_j^{-A+1} \Delta m = 1 M_\odot. \tag{2}$$

It is taken that the initial mass function is described by the Salpeter law with A=2.35 [15]. The upper and lower mass limits of formed stars are taken to equal $m_U = 120 M_\odot$ and $M_L = 0.1 M_\odot$ accordingly. The mass of matter which has been ejected by the stellar population at the moment t is

$$Q_m(t) = \int_0^t \int_{m_L}^{m_U} \dot{Q}(m,\tau) \phi_0 m^{-A} dm d\tau, \tag{3}$$

where $\dot{Q}(m,\tau)$ is rate of mass loss by a star with mass m and lifetime $\tau$. The mass of a synthesized element i ejected by all stars of a population at the moment t:

$$Q_i(t) = \int_0^t \int_{m_L}^{m_U} \dot{Q}(m,\tau)(Z_i(m,\tau) - Z_i(0)) \phi_0 m^{-A} dm d\tau, \tag{4}$$

where $Z_i(0)$ is abundance of element i in gas from which the stars were formed, $Z_i(m,\tau)$ is abundance of element i in matter ejected by stars with mass m and lifetime $\tau$ [2].

We assume that each fragment evolves as a closed system and that its evolution is considered within the framework of simple model of chemical evolution of galaxies. Star formation process in a fragment is considered as sequence of bursts with a population of stars is formed during each burst. The mass of gas $m_g$, the mass of element i $m_i$ and the mass converted into stellar remains $m_s$ at the start of the star formation burst $t_{b_j}$:

$$m_g(t_{b_j}) = m_g(t_{b_{j-1}}) - m_{b_{j-1}} - \sum_{k=1}^{j-1} m_{b_k} [Q_m(\tau_{j,k}) - Q_m(\tau_{j-1,k})] \tag{5}$$

$$m_i(t_{b_j}) = m_i(t_{b_{j-1}}) - m_{b_{j-1}} z_i(t_{b_{j-1}}) - \\ - \sum_{k=1}^{j-1} m_{b_k} [Q_m(\tau_{j,k}) - Q_m(\tau_{j-1,k})] z_i(t_{b_k}) + \\ + \sum_{k=1}^{j-1} m_{b_k} [Q_i(\tau_{j,k}) - Q_i(\tau_{j-1,k})] \tag{6}$$

$$m_s(t_{b_j}) = m_s(t_{b_{j-1}}) + m_{b_{j-1}} - \sum_{k=1}^{j-1} m_{b_k} [Q_m(\tau_{j,k}) - Q_m(\tau_{j-1,k})] \tag{7}$$

where

$$\tau_{j,k} = t_{b_j} - t_{b_k}$$
$$\tau_{j-1,k} = t_{b_{j-1}} - t_{b_k},$$

where $m_{b_j}$ is the mass of a star formation burst j , $z_i(t_{b_j})$ is the abundance of element i at the moment $t_{b_j}$, $\tau_{j,k}$ is the age of burst k at the moment $t_{b_j}$[2]. The second part on the right hand side of equation 6 takes into account a mass change of element i in the interstellar medium as a result of conversation of a gas into stars. The first sum on the right hand side of equation 6 describes the contribution of matter ejected by stars as if the stars would eject unconverted matter. The second sum represents the contribution of syntesized elements. Having solved numerically the equations 5 – 7 we obtain a metallicity distribution function for the stars formed in a fragment of unit mass (with 1 $M_\odot$) with an astration level $s = 1$, see Fig. 1a. Since the evolution of all fragments is described by the simple model the metallicity distribution function of stars in all fragments will have same shape (Fig. 1a) but the upper metallicity limit of the stars in fragments with different astration levels will be different.

In order to test the validity of the building-up numerical model we carried out a comparison of numerical results with analytical. The comparison of stellar metallicity distributions calculated within the framework of the numerical model

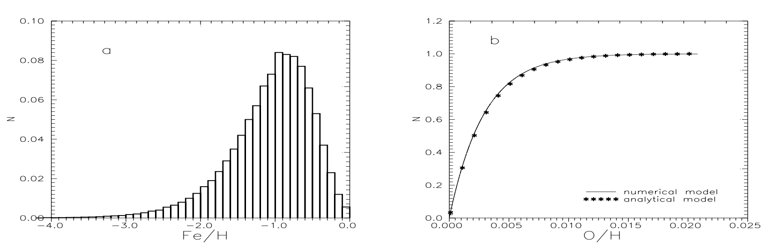



**Fig. 1.** (**a**) the metallicity distribution function calculated in the framework of the numerical model for a fragment with a level astration $s = 1$. (**b**) A comparison of the results from numerical (*solid line*) and analytical (*asterisks*) modelling of the cumulative stellar metallicity distribution

and with the help of the analytical expression of the simple model taken from [3] is shown in Fig. 1b.

Let us now consider the method by which it is possible to obtain masses of fragments where the halo stellar population was formed. Let $a_j$ be the fraction of stars found in a range of metallicities $Z_j$, $Z_j + \Delta Z$ in a fragment with mass m (in our case $m = 1M_{sun}$). Then an observed amount of halo stars finding in a range of metallicities $Z_j$, $Z_j + \Delta Z$ will be represent the total amount of stars $N_{Z_j} = a_j \sum_{i=1}^{n} m_i$ of a given metallicity $Z_j$ from all fragments whose maximum metallicity exceeds $Z_j$. For given set of metallicities we have

$$
\begin{aligned}
a_1 \cdot m_1 + a_1 \cdot m_2 + a_1 \cdot m_3 + \ldots\ + a_1 \cdot m_n &= N_{Z_1} \\
a_2 \cdot m_2 + a_2 \cdot m_3 + \ldots\ + a_2 \cdot m_n &= N_{Z_2} \\
\vdots \quad &\ \ \vdots \\
a_n \cdot m_n &= N_{Z_n}
\end{aligned}
$$

The fragment with the highest astration level s determines the number of stars with the greatest value of a metallicity $Z_n$. Having solved the set of equations (8a) it is possible to obtain the masses of fragments with maximum metallicities of stars from $Z_1$ up to $Z_n$ (i.e. the number of fragments with unit mass falling on given $Z_j$ and responsible for the halo stars with such metallicity). Thus, using an observed metallicity distribution of a halo stars and a modelled metallicity distribution function for a fragment of unit mass we shall obtain a value of the total mass of the unit mass fragments evolved up to each given value of astration level s. It is necessary to note that we can obtain the total mass of fragments evolved up to a given value of metallicity $Z_j$ but not the number of fragments that are included in this total mass.

## 3 Results and Discussion

### 3.1 The Halo Field Stars of Our and Some Other Galaxies

Let us now compare the obtained set of fragments for the halo field stars of our Galaxy (Fig. 2a), M31 (Fig. 2c), NGC 5128 (Fig. 2e). It is necessary to mention that obtained values of masses of fragments are conditional since in order to obtain the precise masses of fragments the mass of the stellar halos of the investigated galaxies were necessary. This value was only available for our Galaxy. Therefore the value of a halo mass $5{*}10^{10} M_{sun}$ for a halo field stars and $10^9 M_{sun}$ for a halo globular clusters was accepted for all galaxies. The obtained results for our Galaxy show that the halo stellar population was formed in fragments with low astration levels and high metallicity dispersion. The theoretical distribution obtained from a mixture of stars from different fragments of the set quite well reproduces the observable distribution (Fig. 2b). It allows us to consider that the halo field stars of our Galaxy were formed in a fragments with a low astration levels and high metallicity dispersion.

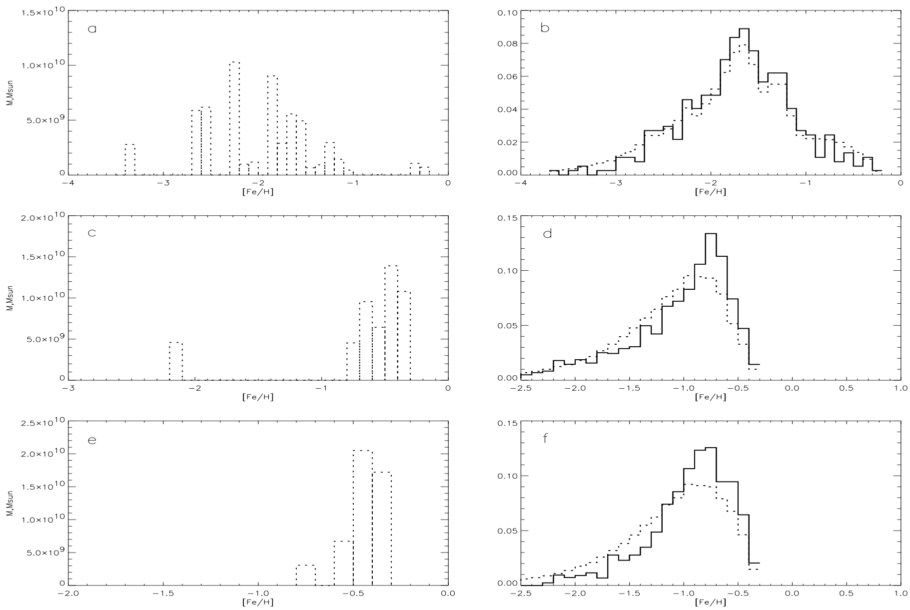



**Fig. 2. a)** The set of fragments obtained from observable metallicity distribution of a halo field stars of our Galaxy **b)** The comparison of observable (*solid line*)[14] and theoretical (*dotted line*) metallicity distributions obtained from a mixture of the stellar population of fragments in a Fig. 2 **c)** the set of fragments obtained from the observable metallicity distribution of halo field stars of M31 **d)** The comparison of the observable (*solid line*)[8] and the theoretical (*dotted line*) metallicity distributions of M31 halo field stars **e)** The set of fragments obtained from observable metallicity distribution of a halo field stars of NGC 5128 **f)** The comparison of the observed (*solid line*)[11] and the theoretical (*dotted line*) metallicity distributions of NGC 5128 halo field stars

Although the metallicity distribution of stars from the obtained set of fragments for M31 reproduces a general wiev of observed distributions it shows more metal-poor stars (Fig. 2d) than it is really observed. The similar pattern is also gained for NGC 5128 (Fig. 2e, 2f). Harris and Harris[12] who investigated the observed metallicity distributions of halo stars of these galaxies concluded that the halo of NGC 5128 and M31 (opposite to the halo of our Galaxy) was formed by the merger of large satellites (similar to LMC, Small Magellanic Cloud and M32 by sizes) rather than by an accretion of smaller stellar systems. Bekki with collaborators[5] have considered a formation of an ellipticals (in particular NGC 5128) by merging of a spiral galaxies and they also have found out that the stellar halo of elliptic galaxies formed by this way, is populated mainly by stars with rather high metallicity ($[m/H] \sim -0.4$) which came from a exterior parts of disks of merging spiral galaxies.It is interesting that if a halo mass fraction of their merging spirals is more then 0.2 the simulated metallicity distribution function shows more metal-poor stars than observable metallicity distribution function. As suggests [5], similarity between distributions of NGC 5128 and of

M31 can be explained in the case when bulge of M31 also was formed by merging of two spirals. Thus, it is possible to assume that if the halo field stars of M31 and NGC 5128 were really formed in mergers the closed box approximation of accreted fragment does not suit this case.

### 3.2 A Globular Cluster System of Our and Some Other Galaxies

The part of a globular cluster metallicity distribution which includes halo globular clusters was only used for computation of masses of fragments. The distribution of halo globular clusters of galaxies was obtained by Gauss approximation of observed distributions in all cases except our Galaxy. In Fig. 3a the set of fragments obtained from observable metallicity distribution of halo globular clusters of our Galaxy is shown. The comparison of observable distribution and distributions obtained from a mixture of stars of fragments in Fig. 3a shows (Fig. 3b) that the obtained set of fragments not so well reproduces observable metallicity distribution of halo globular clusters of our Galaxy as it was expected. Although metallicity distribution for halo globular clusters is more narrow than for field stars (Fig. 3) peaks of observable metallicity distributions of halo field stars of and halo globular clusters of our Galaxy coincide. The Fig. 3d shows age - metallicity relation for globular clusters of our Galaxy. The comparison of observations with tracks of fragments of a unit mass (evolved as closed system) shows that the stellar population of fragments whose star formation began in a different times can reproduce declination and scatter of observable values of ages and metallicities in Fig. 3d. Thus it is quite possible that the halo globular clusters could be formed together with a major part of halo field stars in the same fragments and consideration of formation of halo globular clusters of our Galaxy as an isolated subsystem is not meaningful.

The comparison of observed and theoretical (obtained from a mixture of stars of the fragments (Fig. 3e)) distributions shows (Fig. 3f) that the theoretical distribution don't well reproduce the metallicity distribution of halo globular clusters of M31. Some surplus of metal-poor stellar population takes place here. The peaks of observed metallicity distributions of halo field stars and of halo globular clusters of M31 do not coincide (the observations give $[Fe/H]$ $\sim$-0.6 [6] and $[Fe/H]$ $\sim$-1.4[4] accordingly). Peak of distribution of halo field stars of M31 coincides with peak of distribution of bulge globular clusters which value [Fe/H] also makes $\sim -0.6$ [4]. Apparently, in the case of M31 we really deal with the population of a spheroid, as it was already mentioned in [8]. In such case, it is possible to consider the evolution of halo globular clusters isolately. However, in any case the surplus of the metal-poor stellar population in theoretical distribution of M31 halo globular clusters (as well as halo field stars) does not allow to assume that the halo globular clusters were formed in fragments evolved as the closed system.

The theoretical metallicity distribution o NGC 5128 (Fig. 3h) which was obtained from a mixture of the stellar population of fragments in a Fig. 3g shows the surplus of the metal- poor stellar population.

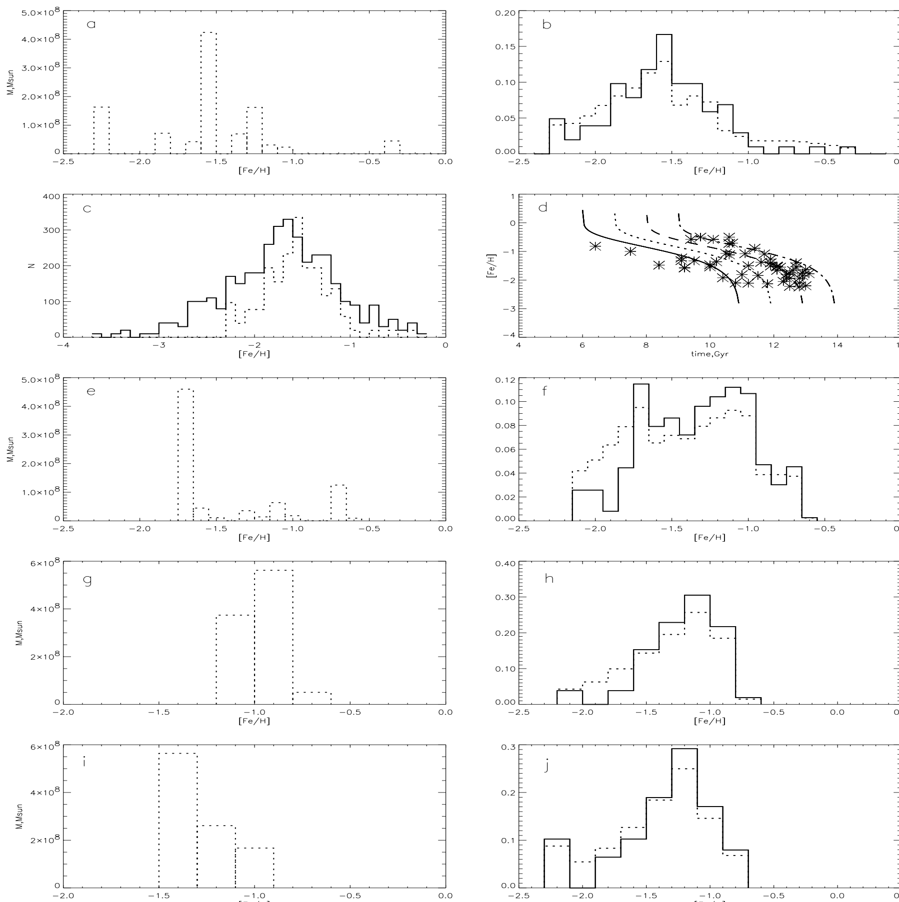


**Fig. 3.** **a)** The set of fragments obtained from the observed halo globular cluster metallicity distribution of our Galaxy **b)** The comparison of observable (*solid line*)[1] and theoretical (*dotted line*) metallicity distribution of Galaxy halo globular clusters. **c)** The comparison of observed metallicity distributions of halo field stars (*solid line*)[14] and halo globular clusters (*dotted line*) [1] for our Galaxy **d)** Age - metallicity relation for globular clusters of our Galaxy. The evolutionary tracks of unit mass fragments (with the time of evolution of 5 Gyr) whose star formation began 11, 12, 13 and 14 Gyr ago is shown by lines. Observational data are taken from [16, 17] **e)** The set of fragments obtained from observable metallicity distribution of halo globular clusters of M31 **f)** The comparison of observable (*solid line*)[4] and theoretical (*dotted line*) metallicity distributions of M31 halo globular clusters. **g)** The set of fragments obtained from observable metallicity distribution of halo globular clusters of NGC 5128 **h)** The comparison of observable (*solid line*)[10] and theoretical (*dotted line*) metallicity distribution of NGC 5128 halo globular clusters **i)** The set of fragments obtained from observable metallicity distribution of halo globular clusters of M87 **j)** The comparison of observable (*solid line*)[7] and theoretical (*dotted line*) metallicity distribution of M87 halo globular clusters

The comparison of metallicity distributions of halo field stars and halo globular clusters of NGC 5128 shows an discrepancy of distribution peaks (the observations give $[Fe/H]$ ∼-0.75 [13] and $[Fe/H]$ ∼-1.11[9] accordingly).

The theoretical distribution for a case M87 (Fig. 3j) generally reproduces an overall view of observed metallicity distribution though the obtained number of globular clusters in each bin is not corresponding with the number of globular clusters in observed metallicity distribution of halo globular clusters for M87.

## 4 Conclusions

Our results allow us to conclude that halo field star formation in our Galaxy can be occured in the fragments evolved as closed system. If the formation of halo field stars of M31 and NGC 5128 has occured by mergers of massive fragments the closed box model of merging fragments doesn't suit this case. The formation of halo field stars of these galaxies perhaps is not associated with the formation of the halo globular clusters in these galaxies In our Galaxy the formation of the halo field stars could be associated with the halo globular cluster formation. The formation of halo globular cluster subsystem as an isolated subsystem in ellipticals (NGC 5128 and M87) is more probable then in spirals (Galaxy and M31). If the halo globular clusters were formed in fragments, a consideration of fragment's evolution as an closed system is not enough to reproduce the observed metallicity distribution of halo globular clusters of ellipticals as spirals.

The autor would like to thank the Organizing Commitee for their kind hospitality and partial financial support. The autor also thanks Pilyugin L.S. and the anonymous referee for a series of valuable notes and help in process of preparation of the paper, Khomenko E.V. for her invaluable help in the development of the necessary programs, Dr.Cardwell and Ms.Rebecca from the IAC for the reading and correcting of this paper. This work was supported by the Ukrainian Fund of Fundamental Researches (grant 02.07/00132)

## References

1. T.V. Borkova, V.A. Marsakov: AZh **77**, 750 (2000)
2. L.S. Pilyugin: AZh **71**, 825 (1994)
3. R.J. Tayler: *Galaxies: Structure and Evolution* (Mir, Moscow 1981)
4. P. Barmby, J.P. Huchra, J.P. Brodie, D.A. Forbes, L.L. Schroder, C.L. Grilmair: Astron. J. **119**, 727 (2000)
5. K. Bekki, W.E. Harris, G.L.H. Harris : astro-ph/0212545
6. M. Bellazzini, C.Cacciari, L.Federici, F. Fusi Pecci, M. Rich:astro-ph/0212531
7. J.G. Cohen, J.P. Blakeslee, A. Ryzhov: Astrophys. J. **496**, 808 (1998)
8. P.R. Durrell, W.E. Harris, C.J. Pritchet: Astron. J. **121**, 2557 (2001)
9. H.Eerik, P.Tenjes:astro-ph/0212522
10. G. Harris, D. Geisler, H. Harris, J. Hesser: Astron. J. **104**, 613 (1992)
11. G. Harris, W. Harris, G. Poole: Astron. J. **117**, 855 (1999)
12. G. Harris, W. Harris: Astron. J. **120**, 2423 (2000)

13. W.E.Harris, G.L.H. Harris: Astron.J. **122**, 3065 (2001)
14. S.G. Ryan, J.E. Norris: Astron. J. **101**, 1865 (1991)
15. E. Salpeter: Astrophys. J. **121**, 161 (1955)
16. M.Salaris, A. Weiss: astro-ph/9704238
17. M. Salaris, A.Weiss A.: astro-ph/0204410
18. L. Searle, W. Zinn: Astroph. J. **225**, 357 (1978)

# Model of Ejection of Matter from Dense Stellar Cluster and Chaotic Motion of Gravitating Shells

Maxim V. Barkov[1], Vladimir A. Belinski[2],
Genadii S. Bisnovatyi-Kogan[1], and Anatoly I. Neishtadt[1]

[1] Space Research Institute, Russian Academy of Sciences, 117997, 84/32 Profsoyuznaya Str, Moscow, Russia;
[2] National Institute of Nuclear physics (INFN) and International Center of Relativistic Astrophysics (ICRA), Rome, Italy

**Abstract.** It is shown that during the motion of two initially gravitationally bound spherical shells, consisting of point particles moving along ballistic trajectories, one of the shell may be expelled to infinity at subrelativistic speed $v_{exp} \leq 0.25c$. The problem is solved in Newtonian gravity. Motion of two intersecting shells in the case when they do not runaway shows a chaotic behaviour. We hope that this simple toy model can give nevertheless a qualitative idea on the nature of the mechanism of matter outbursts from the dense stellar clusters.

## 1 Introduction

Dynamical processes around supermassive black holes in quasars, blazars and active galactic nuclei (AGN) are characterised by violent phenomena, leading to formation of jets and other outbursts. Here we consider the possibility of a shell outburst from a supermassive black holes (SBH) surrounded by a dense massive stellar cluster, basing on a pure ballistic interaction of gravitating shells oscillating around SBH.

Investigation of spherical stellar clusters using shell approximation was started by Hénon [6], and than have been successfully applied for investigation of the stability [7], violent relaxation and collapse [5, 6, 10], leading to formation of a stationary cluster. Investigation of the evolution of spherical stellar cluster with account of different physical processes was done on the base of a shell model in the classical series of papers of L. Spitzer and his coauthors [15–22], see also [12].

Numerical calculations of a collapse of stellar clusters in a shell approximation [3, 4, 23] had shown, that even if all shells are initially gravitationally bound, after a number of intersections some shells obtain sufficient energy to become unbound, and to be thrown to the infinity. In the Newtonian gravity the remnant is formed as a stationary stellar cluster, and in general relativity SBH may be formed as a remnant.

Important example of a quasi-spherical mass ejection is a relativistic collapse of a spherical stellar system, which is considered [8, 9, 14, 24] as the main mechanism of a formation of supermassive black holes in the galactic centers.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 357–364, 2003.

Approximation of such collapse by consideration of spherical shells is the simplest approach, which reflects all important features of such collapse [3–5]. Our consideration is related to the motion of stars (shells) which remain outside the newly formed supermassive black hole.

Here we consider a simplified problem of a motion of two massive spherical shells, each consisting of stars with the same specific angular momentum and energies, around SBH. In a more complicated case of numerous shell intersections this elementary act is a key process of the energy exchange between stars and of matter ejection. This is also the elementary process leading to the violent relaxation of the cluster [10], studied in the shell approximation in [3–5].

Development of chaos during the motion of two gravitating intersecting shells had been found first by Miller and Youngkins [11] in the oversimplified case with a pure radial motion and reflecting inner boundary. We have found [2] a chaotic behaviour in a more realistic model where stars with the same energy and angular momentum are dispersed isotropically over the spherical shell, and each star moves along its ballistic trajectory in the averaged gravitational field, with account of the shell self-gravity. We find conditions at which one of two shells is expelling to infinity taking energy from another shell. We find a maximum of the velocity of the outbursting shell as a function of the ratio of its mass $m$ to the mass $M$ of SBH using Newtonian theory of the shell's motion.

We show, that for equal masses of two shells the expelling velocity reach the value $v_{max} \approx 0.3547 v_p$ at $m/M = 1.0$ , and $v_{max} \geq 0.3 v_p$ was obtained at the masses of shells $0.25 \div 1.5M$, where the parabolic velocity of a shell in the point of a smallest distance to the black hole $v_p$ may be of a considerable part of $c$.

In Sects. 2,3 we describe outburst effect and in Sects. 4 we present the evidence of the chaos in the system of intersecting shells. The exact solution of these problems in the context of General Relativity have been found in [1].

## 2 Two Shells Around SBH

Physically the nature of the ballistic ejection is based on the following four subsequent events. The outer shell is accelerated moving to the center in a strong gravitational field of a central body and inner shell. Somewhere near the inner minimal radius of the trajectory shells intersect. After that the former outer shell is decelerating moving from the center in the weaker gravitational field of only one central body. The second intersection happens somewhere far from the center. That may result in the situation when the total energy (negative) of the initially gravitationally bound outer shell is becoming positive as a result of two subsequent intersections with another shell. The quantitative analysis of this process is done in [2].

Equation of motion of a shell with mass $m$ and total conserved energy $E$ in the field of a central body with mass $M$ is

$$E = \frac{mv^2}{2} - \frac{Gm(M + m/2)}{r} + \frac{J^2 m}{2r^2}, \tag{1}$$

where $v = dr/dt$ is the radial velocity of the shell and $J^2m/2r^2$ is the total kinetic energy of tangential motions of all particles, which the shell is made up from. The constant $J > 0$ has that interpretation that $Jm$ is the sum of the absolute values of the angular momenta of all particles.The term $m/2$ in (1) is due to the self-gravity of the shell.

Let us consider two shells with parameters $m_1, J_1$ and $m_2, J_2$ moving around SBH with mass $M$. Let the shell "1" be initially outer and the shell "2" be the inner one. Then equations of motion are:

$$E_{1(0)} = \frac{m_1 v_{1(0)}^2}{2} - \frac{Gm_1(M + m_1/2 + m_2)}{r} + \frac{J_1^2 m_1}{2r^2}, \tag{2}$$

$$E_{2(0)} = \frac{m_2 v_{2(0)}^2}{2} - \frac{Gm_2(M + m_2/2)}{r} + \frac{J_2^2 m_2}{2r^2}. \tag{3}$$

By the index (0) we mark the initial evolution stage before the first intersection of the shells. Assume that both shells are moving to the center. Such shells intersect each other at a some radius $r = a_1$ and at some time $t = t_1$ after first intersection the shell "1" becomes inner and shell "2" outer and at a some radius $r = a_2$ and at some time $t = t_2$ they have second intersection. The motion of the shells after first intersection is designated by the index (1), and after second intersection is designated by the index (2):

$$E_{1(2)} = E_{1(1)} - \frac{Gm_1 m_2}{a_2} = E_{1(0)} + Gm_1 m_2 \left( \frac{1}{a_1} - \frac{1}{a_2} \right), \tag{4}$$

$$E_{2(2)} = E_{2(1)} + \frac{Gm_1 m_2}{a_2} = E_{2(0)} - Gm_1 m_2 \left( \frac{1}{a_1} - \frac{1}{a_2} \right). \tag{5}$$

Let us describe the situation, when one shell is ejected to infinity after intersection of to initially bound shells. We consider a case when $a_2$ is larger then $a_1$ so, that the second term in (4) has larger absolute value, then the first one, the first shell gains a positive energy and goes to infinity. Both shells have initial negative energies $E_{1(0)}$ and $E_{2(0)}$, but with small enough absolute values. The first shell takes the energy from the second one, which is becoming more bound with larger absolute value of the negative energy $E_{2(2)}$, according to (4).

## 3 Numerical Solution

Let us illustrate the foregoing scenario by an exact particular example of two shells of equal masses. We choose parameters in the following way:

$$m_1 = m_2 = m, \qquad E_{1(0)} = E_{2(0)} = 0, \qquad J_1 < J_2, \tag{6}$$

In fact such exact solution represents the first approximation to the more general situation when $E_{1(0)}$ and $E_{2(0)}$ are non-zero (negative) but small in that sense that both modulus $|E_{1(0)}|$ and $|E_{2(0)}|$ are much less than $Gm^2/a_1$.

We assume that initially both shells are moving towards the center. It follows from (2), (3) that under condition (6) such shells will intersect inescapably at the point $r = a_1,\ t = t_1$. After the second intersection at $r = a_2,\ t = t_2$ the shell "1" will be thrown to infinity with expelling velocity $v_{1exp}$. It follows from (4) that

$$v_{1exp} = v_{1(2)}\big|_{r\to\infty} = \sqrt{2Gm\left(\frac{1}{a_1} - \frac{1}{a_2}\right)}. \tag{7}$$

In order to construct a solution with maximal possible $v_{1exp}$ we consider a case when initially inner shell "2" reaches the inner turning point at minimal possible radius $r = r_m$ of the order of few $r_g = 2GM/c^2$, and intersects with the initially outer shell after, during its outward motion (see Fig. 1).

Between the first and second intersection there exists the inner turning point of the shell "1" (now inner shell). We take that additional restriction that the shell "1" reaches this turning point also at the minimal possible radius $r = r_m$. It is easy to show that a wide class of solutions with such restriction really exists (see Fig. 1).

We now introduce the "parabolic" velocity $v_p$ of the any outer shell at the point of its minimal distance to the center $r = r_m$ as:

$$v_p = \sqrt{\frac{2G(M + 3m/2)}{r_m}}. \tag{8}$$

Then we have

$$\frac{v_{1exp}}{v_p} = \sqrt{\frac{m}{M + 3m/2}\left(\frac{r_m}{a_1} - \frac{r_m}{a_2}\right)}. \tag{9}$$

Consequently this ratio is also a function of $a_1/r_m$ and $m/M$. We find numerically the value of $a_1/r_m$ maximizing the ratio $v_{1exp}/v_p$ which value is a function

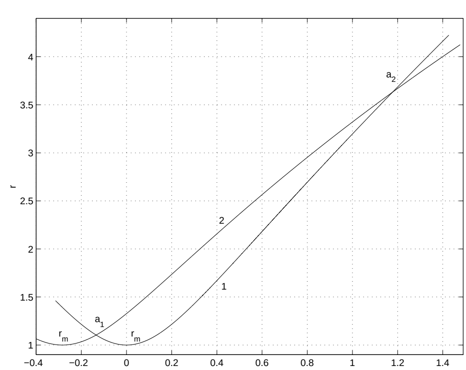



**Fig. 1.** Time dependence of radii r of two shells (in units $r_m$) on time t (in units $r_m/v_p,\ v_p = \sqrt{2G(M + 3m/2)/r_m}$). The shells intersect at points $a_1 = 1.1401$, $a_2 = 3.6316$. Here $m/M = 0.1$, and after the second intersection the first shell is running away with velocity at infinity $v_{1exp} = 0.23415 v_p$.

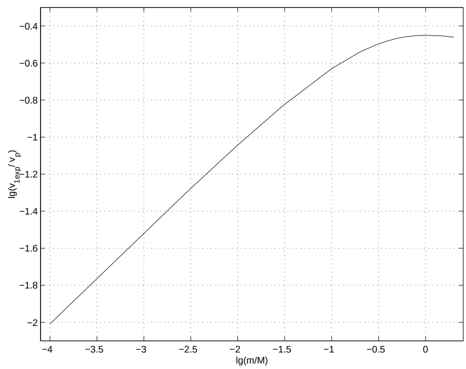



**Fig. 2.** Dependence of the maximum runaway velocity $v_{1exp}$ in units $v_p$ on the ratio of the shell mass to the mass of central body $m/M$. The maximum value of $v_{1exp} = 0.3547$ corresponds to $m/M = 1.0$.

of $m/M$ only. Dependencies for $v_{1exp}/v_p$ as functions of the parameter $m/M$ for these maximizing solutions are given in Fig. 2.

The numerical calculations (illustrated by the Fig. 2) shows that the runaway velocity $v_{1exp}$ reaches its maximal possible value at $m/M \approx 1.0$ and it is $v_{1exp} \approx 0.3547 v_p$. If we consider the shells around SBH then the minimal radius $r_m$ of the shell "1" orbit cannot be less then the two gravitational radii $2r_g = 4G(M + 3m/2)/c^2$. In the extreme case when $r_m = 2r_g$ we have $v_p \sim c/\sqrt{2}$ and for the maximal possible runaway velocity we get $v_{1exp} \sim 0.25c$.

## 4 Chaos in the Shell Motion

The first evidence that the motion of two intersecting shells can show chaotic character was given by B.N. Miller & V.P. Youngkins [11]. They investigated the special case when the central body is absent and particles making up the shells are moving only in radial direction (in our notation $M = 0$ and $J_1 = J_2 = 0$) with an artificial reflection at a given inner radius. This situation, however, cannot model astrophysical cluster with massive nuclei, and also the problem of the influence of central Newtonian singularity arises, which need some additional care. In any case a study of more physically realistic models with nonzero $M, J_1$ and $J_2$ from the point of view of possible chaotic behaviour represents an essential interest. We report here some numerical results for such more general two-shells model which were investigated in the previous sections but again for the shells with equal masses.

We fix here the initial specific angular momenta and energies, different for both shells, initial radii of the shells, and vary only the mass ratios of the shells and the central body. It was found that at our choice of parameters the motion of the shells becomes chaotic at $m/M = 2\%$.

For more clear understanding of the problem let us consider the following simplification. Put one of the shell mass equal to zero. We fix the specific angular momenta of the shells (the angular momentum of the light shell is less than the angular momentum of the heavy one) and initial specific energy of the heavy shell. The initial specific energy of the light shell is chosen to satisfy the following condition: specific energy of the heavy shell is less than that of the light one. Due to these conditions on every turnover the light shell has two intersections with the heavy shell.

We implement the approach from [13] for analysis of this model. Namely, we study Poincaré section and Poincaré return map of the problem. We define in the phase space the surface for Poincaré section by conditions that the radial velocity of the light shell is 0, and the radius of the light shell has a local minimum. Let $E, E'$ and $\eta, \eta'$ (defined modulus $2\pi$) be specific energies of the light shell and phases of the heavy shell at two subsequent time moments when the radius of the light shell has minima. (Phase of the heavy shell is just the mean anomaly of Keplerian motion of this shell). The Poincaré map sends pair $E, \eta$ to the $E', \eta'$. This map is area-preserving. The trajectories of several initial points under the action of the Poincaré map are shown in Fig. 3. While for the problem under consideration the forces are discontinuous, the Poincaré return map is analytic, and therefore KAM theory can be applied for small values of $m/M$, where $m = m_2$. Note that $E - E' \sim m/M$, $\eta - \eta' \sim (-E')^{-3/2}$, similarly to [13]. Following [13], we get the estimate for the border value $E^*$ of the chaotic region in E-space (see Fig. 3) $E^* \sim (m/M)^{2/5}$, i.e. the chaotic behavior is possible at any value of $m/M$ (see Figs. 3-4). Our numerical simulation suggests relation $E_1^* \sim (m/M)^{0.42}$; at $E > E_1^*$ runaway of the light shell is possible and at $E < E_1^*$ the light shell is not be able to escape. The region $E < E_1^*$ up to a residue of a small measure is filled with invariant curves of the Poincaré map in accordance with KAM theory. Another relation $E_2^* \sim (m/M)^{0.46}$ was found for energy value separating the region $E > E_2^*$ in where the shell motion is chaotic and the region ($E_1^* < E < E_2^*$) in which the regular motion is possible. As one can see, the numerical simulations and analytical estimation give close results.

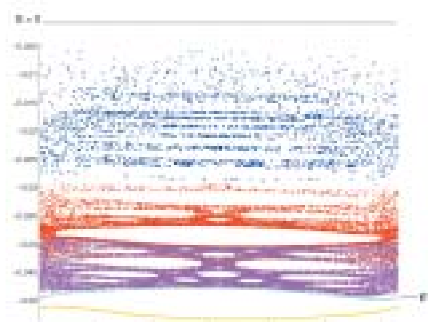


**Fig. 3.** Distribution of points in (E,$\eta$)-space for $m/M = 0.005$. The upper line indicates the border of capture of a shell, i.e. E=0. $E^*$ is the lower limit of the chaotic motion, related to the breakdown of the invariant curve in the KAM theory. The structure is similar [13] for different values of the $m/M$.

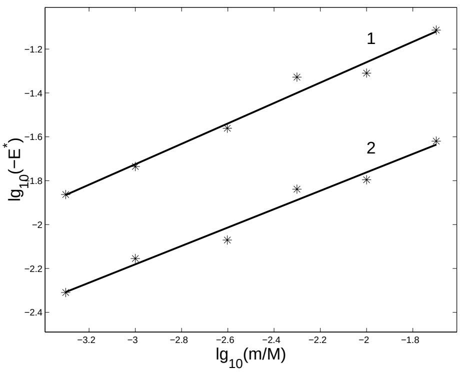



**Fig. 4.** The lines divide the plane into 3 regions. Above the first line ($E_1^* \sim -(m/M)^{0.42}$) escape to infinity is impossible for any initial parameter values. Below the second line ($E_2^* \sim -(m/M)^{0.46}$) escape to infinity is possible for any initial parameter values. For region between lines "1" and "2" the possibility of escaping to infinity depends on initial conditions.

In the case when the light shell does not have two intersections on every turnover, the Poincaré return map is continuous but not smooth. In this case KAM theory is not applicable. Invariant sets become belt-like and have non zero width.

## Acknowledgement

The work of G.S.B.-K. and M.V.B. was partly supported by RFBR grant 99-02-18180, and INTAS grant 00-491. The work of A.I.N. was partly supported by RFBR grant 03-01-00158, and INTAS grant 00-221.

# References

1. Barkov M. V., Belinski V. A., and Bisnovatyi-Kogan G. S., 2002, JETP, 95, 3, 371–391.
2. Barkov M. V., Belinski V. A. and Bisnovatyi-Kogan G. S.,2002, MNRAS 334, 338–344
3. Bisnovatyi-Kogan, G. S., Yangurazova, L. R. 1987, Asrtofizika 27, 79
4. Bisnovatyi-Kogan, G. S., Yangurazova, L. R. 1988, Ap. Sp. Sci. 147, 121
5. Gott, J. P. 1975, ApJ 201, 296
6. Hénon, M. 1964, Ann. d'Ap. 27, 83
7. Hénon, M. 1973, Astron. Ap. 24, 229
8. Ipser, J. . 1980, ApJ 238, 1101
9. Lightman, A. P., Shapiro, S. L. 1978, Rev. Mod. Phys. 50, 437
10. Lynden-Bell, D. 1967, MNRAS 136, 101
11. Miller B. and Youngkins V. 1997, Chaos 7, 187.
12. Palmer P. and Voglis N., 1983, MNRAS 205, 543.
13. Petrosky T.Y. 1986, Phs. Lett. A, 117, 7, 328-332

14. Rasio, F., Shapiro, S., Teukolsky, S. 1989, ApJ 336, L63
15. Spitzer, L. Jr., Hart, H. M. 1971a, ApJ 164, 399
16. Spitzer, L. Jr., Hart, H. M. 1971b, ApJ 166, 483
17. Spitzer, L. Jr., Shapiro, S. L., 1972, ApJ 173, 529
18. Spitzer, L. Jr., Thuan, T. X. 1972, ApJ 175, 31
19. Spitzer, L. Jr., Chevalier, R., 1973, ApJ 183, 565
20. Spitzer, L. Jr., Shull, J. M., 1975a, ApJ 200, 339
21. Spitzer, L. Jr., Shull, J. M., 1975b, ApJ 201, 773
22. Spitzer, L. Jr., Mathieu, R. D., 1980, ApJ 241, 618
23. Yangurazova, L. R., Bisnovatyi-Kogan, G. S. 1984, Ap. Sp. Sci. 100, 319
24. Zeldovich, Ya. B., Podurets, M. A. 1965, AZh 42, 963 (transl. Soviet Astron. -AJ, 29, 742 [1966])

# Direct vs Merger Mechanism Forming Counterrotating Galaxies

Maria Harsoula and Nikos Voglis

Academy of Athens, Research Center for Astronomy, 14 Anagnostopoulou Str., GR-10673, nvogl@cc.uoa.gr, mharsoul@phys.uoa.gr

**Abstract.** A comparison of the formation of counterrotating elliptical galaxies is made in two alternative scenarios: (a) The scenario of merging of a primary and a satellite galaxy, and (b) the direct scenario, in which counterrotating galaxies are formed directly from cosmological initial conditions. We conclude that, although both scenarios might have worked in parallel, the scenario of merging has a large number of parameters that should be well tuned to form a counterrotating galaxy in contrast to the direct scenario that is controlled by only two or three parameters. This could give more chances to the direct scenario.

## Introduction

It is known, nowadays, from observations, that almost 1/3 of the elliptical galaxies present a counterrotating core (CRCs hereafter) with respect to the rest of the galaxy [4], [14], [15], [5], [22], [23]. A number of cases of spiral counterrotating core galaxies are also reported [8].

The scenarios proposed in the literature explaining the formation of such galaxies are split into two main categories: a) the merger scenarios and b) the non-merger scenarios. According to the first scenarios CRCs ellipticals are the result of the merging between two galaxies. The merging can be either dissipationless, i.e. between a main galaxy and a satellite one [2], [1], or between two spiral galaxies [26], [3]. Similar scenarios consider episodic gas infall, or accretion of a gas rich companion [6], [7]. Most of these scenarios fail to explain the reddening and the increased metallicity of the core of the CRC galaxy.

A non-merger scenario has been proposed by Hau and Thomson [18] according to which tidal torquing on the main body of the galaxy is caused by another passing galaxy. This torque can reverse the sense of rotation of the outer parts of the galaxy and therefore can in general produce a kinematically decoupled core.

An alternative non-merger scenario, is the one that we call direct scenario and it was proposed by Voglis et al [28], [17], [29], where a bar-like density excess in the early Universe (i.e. Decoupling) that makes bound the material of a protogalactic cloud, can work as a seed to form a CRC galaxy. The great advantage of this scenario is that it explains the almost total lack of any population differences between the core and the main body of the galaxy as well as between the CRC galaxies and other normal galaxies.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 365–379, 2003.

A brief discussion of the direct scenario is given in Sect. 1. In Sect. 2 the initial conditions and the results of a numerical experiment are described, in which a counterrotating galaxy is formed from cosmological clumpy initial conditions. In Sect. 3 we describe the initial conditions and the results of another experiment in which a counterrotating galaxy is formed by the merging process. The parameters in the latter case were arranged so that the two galaxies have almost the same rotation velocity curve. This allows a comparison between the two scenarios discussed in Sect. 4.

## 1 Basic Features of the Direct Scenario

Let's consider a density perturbation at the early post-decoupling Universe which has a bar-like shape and is formed by small clumps of mass. The most tightly bounded particles have a quadrupole moment with major axis A1 along a random direction (Fig. 1). Then the distribution of the more loosely bound particles has a quadrupole moment with major axis along the direction A2. This protogalactic cloud is surrounded by an anisotropic environment which creates a tidal field with random orientation of its main axis (direction of larger forces).

If the axis A1 is along the direction of 2nd-4th quadrant, then these particles will acquire positive angular momentum while the more loosely bound particles being mainly along the direction 1st-3rd-quadrant (along A2 in Fig. 1), acquire negative angular momentum.

Therefore, this configuration has a certain distribution of angular momentum along the radius, that can initiate a counterrotating system. After the collapse and the mixing of the particles this initial distribution of the angular momentum can survive under some conditions.

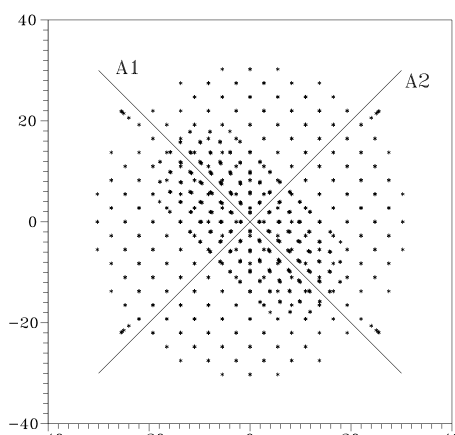



**Fig. 1.** A bar like distribution with major axis A1. Loosely bound particles have major axis A2

The violent relaxation of the system cannot erase all the memory of the initial conditions and therefore the particles remember their initial energies (and as a consequence their initial positions) in a statistical sense, even for a Hubble time.

This scenario is verified by N-body simulations in two different cases of initial conditions, namely quiet initial conditions [28], [17], [16] and clumpy initial conditions [29].

In the case of quiet initial conditions we found that a counterrotating galaxy is formed when the bar-like initial density perturbation has an axial ratio around the value 0.5. If this ratio is changed to 0.4, negative rotation dominates all along the radius, while for a ratio 0.8, positive rotation dominates all along the radius [17].

A second parameter that influences the appearance of counterrotation is the strength of the external tidal field. It was found that a modest external tidal field favors the formation of counterrotating galaxies, while a strong or a weak tidal field favor the formation of galaxies with simple rotation [16].

## 2 Counterrotation from Clumpy Cosmological Initial Conditions

This scenario can be realized in the case of clumpy cosmological initial conditions as follows. We used two different mass scales: the galactic mass scale named G-scale which contains 2 $M_u$ (where $M_u$ is the mass unit, i.e. the mass of a galaxy) and the environment scale named E-scale which contains 664 $M_u$.

The G-scale distribution of particles is composed of a number of $N_g = 5616$ particles of equal mass initially arranged in Lagrangian coordinates $(q_1, q_2, q_3)$ in a cubic grid limited by a sphere. The size of the system is arranged in such a way that when it expands with the Universe it has total energy equal to zero, simulating the expansion of an Einstein-de Sitter Universe (i.e. $\Omega = 1$, all distance increase with time $t$ as $t^{2/3}$).

The positions and velocities of particles, when the system is perturbed, are evolved by the Zeldovich approximation [31], as explained analytically in [29].

The E-scale is resolved into 664 particles put initially in a cubic grid limited by a sphere. Eight central particles of this grid have been removed to make room for the location of the G-scale. Each of the remaining particles of the E-scale has mass equal to 1 $M_u$. Their grid is slightly deformed in such a way that undesirable non-cosmological torques acting of the G-scale due to the discreteness of the E-scale is almost eliminated (see for details [27]). Then the positions and the velocities of the E-scale particles are evolved in such a way so that it simulates correctly the initial strength and the time behavior of the cosmological tidal torque in agreement with the analytical calculations during the linear phase of evolution of the density perturbation. The Y-axis is the axis of strongest forces of the tidal field. In Fig. 2a the overall projection of the initial configuration is shown at a time=0, when the N-body calculations start. The G-scale is shown in magnification in Fig. 2b at the same time.

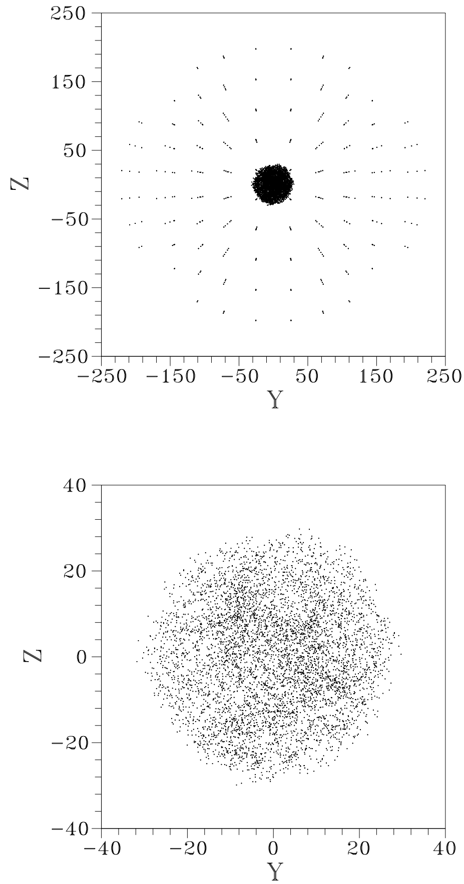



**Fig. 2.** The projection of the initial configuration on the y-z plane at time=0. (**a**) the whole configuration and (**b**) the g-scale only

The evolution of the whole system (of both scales) is followed simultaneously by Aarseth's N-body2 code. The units of time $t_{un}$, length $r_{un}$ and velocity $v_{un}$ in our unit system are expressed in terms of real units as:

$$t_{un} = 0.875\beta^{3/2} Myears \tag{1}$$

$$r_{un} = 1.5\beta(\frac{M_{un}}{M_{12}})^{1/3}(Kpc) \tag{2}$$

$$v_{un} = 1677\beta^{-1/2}(\frac{M_{un}}{M_{12}})^{1/3}(km/sec) \tag{3}$$

where $M_{12} = 10^{12} M_{\odot}$ and $\beta$ is a re-scaling parameter.

The G-scale configuration evolves so that a part of about 4800 particles of it reaches a maximum radius of expansion and then detaches from the general

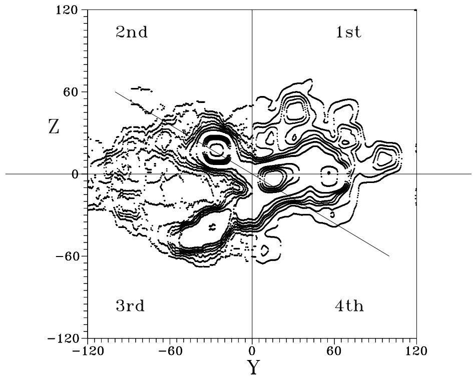



**Fig. 3.** The isodensity contours plotted on the Y-Z plane at time=$400t_{un}$ for the experiment with clumpy initial conditions

expansion. During this period clumps grow inside it, and gradually merge to form a single main clump i.e. a galaxy. 'Violent relaxation' occurs by the process of gradual merging.

For a certain range of the initial parameters i.e., consistent to the values used in the case of quiet initial conditions in Harsoula and Voglis 1998, we detected counterrotation in the relaxed system.

In Fig. 3 the isodensity contours are plotted on the Y-Z plane at a time corresponding to $400t_{un}$. It's obvious that the most tightly bound particles form a bar-like density perturbation with a major axis of about $50^o$ with respect to the tidal axis. These particles acquire therefore positive angular momentum, while the less tightly bound particles have a major axis almost perpendicular to the axis of the most bound ones and acquire negative angular momentum.

Beyond a time of $4000t_{un}$ the bar and a considerable part of the G-scale around the bar has collapsed and been relaxed to a bound system i.e. a galaxy. The secondary infall is no more important.

Figure 4 shows the rotational velocity profile of an experiment with counterrotation. The dots upon the rotation curve correspond to fractions 10%, 20% etc. of the total mass inside the respective radius. We see that about 60% of the bound mass has positive rotation, while the rest of the bound mass has negative rotation. As it was mentioned above the strength of the cosmological tidal field is important for counterrotation to appear. For example, if we repeat this experiment having the G-scale exposed to a tidal field, 30% stronger counterrotation disappears. Angular momentum of negative sign dominates throughout the whole configuration. The corresponding rotational velocity profile is shown in Fig. 5.

Another parameter that could play a role is the initial orientation of the bar formed by the most tightly bound particles. However, the results are not expected to be particularly sensitive on this orientation, unless the bar is nearly parallel or perpendicular to the tidal axis.

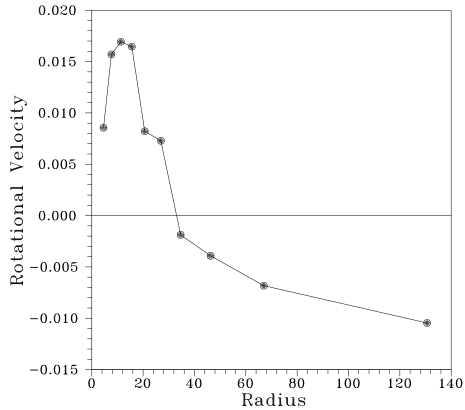



**Fig. 4.** (The rotational velocity profile for an experiment with clumpy cosmological initial conditions giving a counterrotating elliptical galaxy

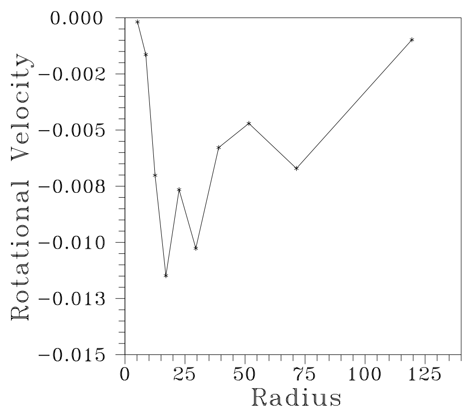



**Fig. 5.** The rotational velocity profile for an experiment with clumpy cosmological initial conditions and a tidal field 30% stronger that the one of Fig. 3

## 3 Counterrotation from Merging with a Satellite Galaxy

Mergers can also produce CRCs. We have already performed some experiments with mergers in order to compare the two scenarios [30]. Here we present a deeper investigation studying also the possibility to distinguish CRCs formed by mergers from those formed by cosmological initial conditions via some observable quantities.

We therefore have performed several experiments of dissipationless merging of a primary galaxy and a satellite galaxy. The primary is designed so that all the orbits rotate in one direction only i.e. contains no retrograde orbits at all.

To form such a primary galaxy, we start with the relaxed configuration of the experiment with the rotational velocity profile of Fig. 5. Then all the velocity components $u_{yz}$, parallel to the Y-Z plane are turned perpendicular to their position vector component $R_{yz}$, with common direction of rotation. Thus the system has no retrograde orbits. It remains in the same virial equilibrium and has a value of the spin parameter $\lambda \approx 0.1$ [25].

The satellite galaxy is produced by isolating a central spherical part of the initial state of the primary galaxy. We have performed three numerical experiments namely A,B and C. In the experiment A the mass of the satellite galaxy is about 20% of the primary's mass, while in the experiment B and C it is about 10% of the primary's mass. In particular, in C the velocities of the particles of the satellite have been multiplied by a factor $0.5^{1/2}$ so that its kinetic energy is reduced by 50% and the satellite becomes quite compact in this experiment. We run the satellite separately for several dynamical times (4000 $t_{un}$) so that it relaxes to a more compact system (with higher central density) before the start of the merging process.

Then, in all three experiments, the satellite galaxy is placed on a retrograde orbit on the plane of rotation of the primary (i.e. on the Y-Z plane) at a distance of $150r_{un}$ in Y-direction and $50r_{un}$ in Z-direction from the center of mass of the primary galaxy. The initial velocity of the center of mass of the satellite galaxy for all the experiments is $U_s \approx Vp/3$ where $Vp$ is the velocity of the parabolic orbit around the primary galaxy.

We notice here that prograde orbits of the satellite galaxy fail to produce counterrotating cores [2]. This is expected for thermodynamical reasons. Furthermore, simulations with orbital inclination different than zero have very little possibility to give counterrotating systems [1]. The reason is that the dynamical friction between the satellite and the rotating primary causes the satellite's orbit to pivot such as to be oriented almost upright in relation to the rotation plane of the primary. A rough idea of the expected change in the orbit's inclination was first given by Chandrasekhar with his dynamical friction formula [10].

In Fig. 6 we can see the initial configuration of the primary and of the satellite galaxy at the start of the N-body run, and the orbit of the center of mass of the satellite galaxy around the common center of mass of the two galaxies for the ex. C. The satellite galaxy describes a spiral orbit around the common center of mass being subject to dynamical friction as it moves towards the primary galaxy. The most compact part of it relaxes at the center.

In Fig. 7 the line with stars shows the rotation velocity of the primary before merging. The line with dots in this figure shows the new form of the rotation velocity of the relaxed system after merging for exp A. This form is similar to the rotational velocity of Fig. 4 obtained from cosmological initial conditions, with a little greater maximum value. It is clear therefore that the satellite sinking inside the primary is able to reverse the rotation near the center, leaving a signature of its proper angular momentum.

In Fig. 8 we present the rotational velocity profiles for all the three experiments. In ex. A (stars) the central region seems to have greater positive values of rotation and this is due to the fact that the satellite has twice the number of particles than in the other two experiments (mass ratio 1:5). Moreover the absolute values of the rotational velocities for the outer parts of the galaxy are smaller than in the other two experiments. This is again due to the mass ratio

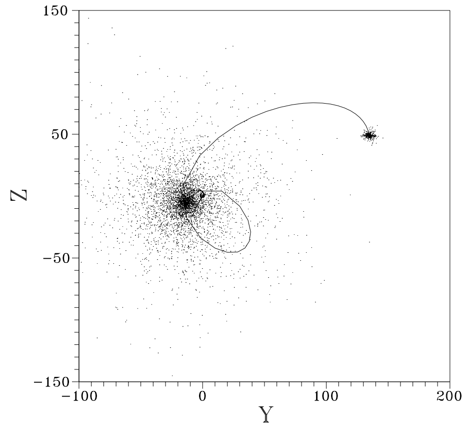



**Fig. 6.** The initial configuration of the primary and the satellite galaxy at time=0 for ex. C and the spiral orbit of the center of mass of the satellite until the whole system is relaxed

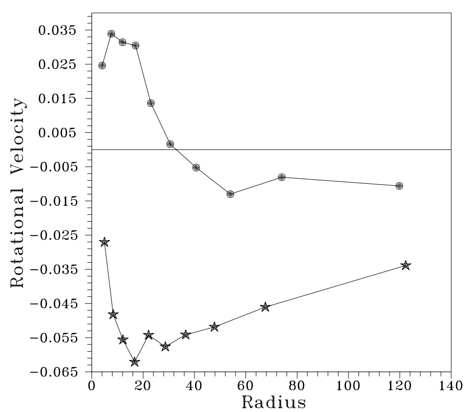



**Fig. 7.** (The rotational velocity profiles, of the counterrotating elliptical galaxy formed by merging of a primary and a satellite galaxy for ex. A (dots) and of the primary galaxy before the merging (stars)

of the two galaxies which is higher is in ex. A (1:5) and therefore the particles of the satellite galaxy, having positive angular momentum with respect to the target galaxy, affect a greater number of particles of the target and can reverse their rotation. In ex. B (dots) where the mass ratio of the two galaxies is 1:10, we observe that the whole galaxy rotates with one sign, except of a small fraction of 10% of the remnant's mass which has a small value of positive rotation. Therefore the mass ratio can seriously affect the rotational velocity profile and the ratio 1:10 is an approximate lower threshold for having counterrotating galaxies. In ex. C (open squares) the mass ratio is still 1:10, but the satellite galaxy is initially more compact and therefore it can preserve better it's orbital angular momentum. Therefore a fraction of 40% of the remnant's mass has positive ro-

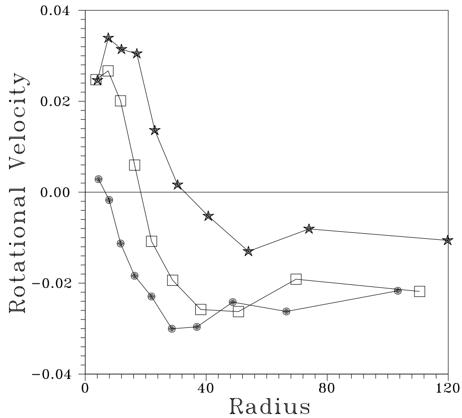



**Fig. 8.** The rotational velocity profile of ex. A (stars), ex. B (dots) and ex. C (open squares) in the case of merger

tation. It seems that in this case the mixing of the particles is less than in the other two cases.

## 4 Comparison of the Results and Discussion

It is useful here to convert the values of the rotational velocities and the radius of Figs. 4 and 7 in real units and compare with the values observed in counterrotating galaxies. Barcells and Quinn [2] proposed a typical counterrotation velocity $40km/sec$ for $M_{un} = 5\text{x}10^{11} M_{\odot}$ and half mass radius $r_o$=$10Kpc$, or at most a factor of 2 higher than this limit. By comparing with our results where $r_o \approx 18 r_{un}$ we get from (2) $\beta \approx 0.42$. From equation (3) we get $v_{un} = 1942 km/sec$. The maximum of the counter rotation velocity of Figs. 3a and 6a is $v_{max} = (0.018\ to\ 0.033) v_{un} \approx\ 35\ to\ 64 km/sec$. Therefore our results, either from cosmological initial conditions or from merging, are in agreement with the observational data.

We have tested the merging scenario for greater values of the initial velocity of the satellite galaxy. We found that for $V_p > U_s > Vp/3$, the final velocity profile shows greater values of positive rotation in the center.

The scenario of merging events and the direct scenario can occur independently. Thus, it is expected that some of the counterrotating galaxies observed today were formed in one way, while others may have formed in the other way. In both cases, our simulations have shown that the distribution of the angular momentum can survive for a Hubble time and therefore it can be observed.

A question that arises is whether, via some observable quantity, counterrotating elliptical galaxies formed by mergers can be distinguished from those formed by cosmological collapses.

Hausman and Ostriker [19] proposed that once a typical victim has a higher central density than the cannibal, its core survives ingestion to produce a core-within-a-core structure. According to Barcells and Quinn [2], if the mass ratio of the merger event is 1 : 5 or higher no obvious trace in the surface brightness profile is left apart from the fact that, in general, there is no flattening in the center. This seems to be confirmed in the simulations of Duncan et al [12] where they have performed mergers with several mass ratios $1 : k$ with $1 \leq k \leq 5$. The surface density profiles of these experiments show no hint of a core-within-a-core structure.

According to our results a hump in the surface density profile is obvious when the mass ratio is 1:10 and the initial reduction of the satellite's initial kinetic energy is greater that 50%. This is seen in Fig. 9 where the surface density profile is plotted for all the three experiments. In Fig. 9a the surface profile of ex. A (solid line) and ex. B (dashed line) is plotted. The two profiles seem quite similar with no obvious hump. The outer part of the remnant galaxy in both experiments

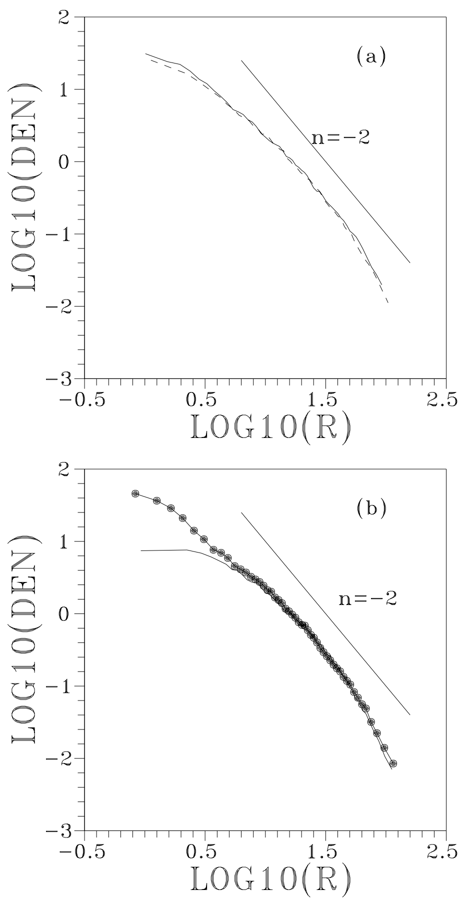



**Fig. 9.** The surface density as a function of the radius in logarithmic scale. (**a**) the solid line corresponds to ex. A and the dashed line to ex. B (**b**) the solid line with dots corresponds to ex. C and the solid line without dots to the primary galaxy before merger

follows approximately Hubbles' power low profile with exponent $n \approx -2$, while it becomes steeper beyond a radius of $\approx 50r_{un}$. In Fig. 9b, the solid line with dots corresponds to the remnant of ex. C and the solid line without dots to the primary galaxy before merging. While the primary profile before merging flattens inside a radius $r \approx 10r_{un}$, the remnant profile continue to rise inward due to the addition of the secondary material and flattens only at the softening length.

In the curve corresponding to ex. C the hump is obvious in the region where the satellite dominates the surface density. A similar density profile is presented, for the first time, in Efstathiou at al [13] for the observed Kinematical decoupled core galaxy NGC5813 and as Kormendy [21] suggested this can be a good candidate for a merger remnant.

In Fig. 10 we compare the surface density profile of ex. C with the one derived from the experiment with cosmological initial conditions. The solid line with dots corresponds to ex. C and the solid line without dots corresponds to the counterrotating system created from cosmological initial conditions. The latter one seems to flatten near the core, while the profile corresponding to the merger continues to rise towards the center, due to the addition of the satellite's material.

A parameter that is always affected from the merging is the galaxy's spin angular momentum. The mixing of particles between the two galaxies and the tidal torques exerted during the merging can cause the reduction of the spin of the primary galaxy, as it was remarked already from 1982 by Negroponte and White [24]. In their simulations the initial spin of the primary galaxy has reduced up to a factor of 3. This is confirmed in Fig. 11 where the spin angular momentum parallel to the x-axis (which is the axis of the rotation of the two

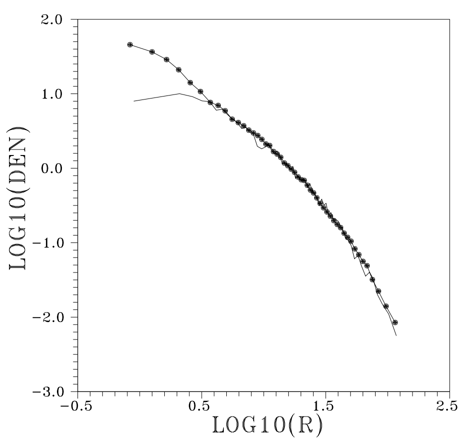



**Fig. 10.** The surface density as a function of the radius in logarithmic scale. The solid line with dots corresponds to the galaxy formed by merger (ex. C) and the solid line (without dots) corresponds to the galaxy formed by clumpy cosmological initial conditions

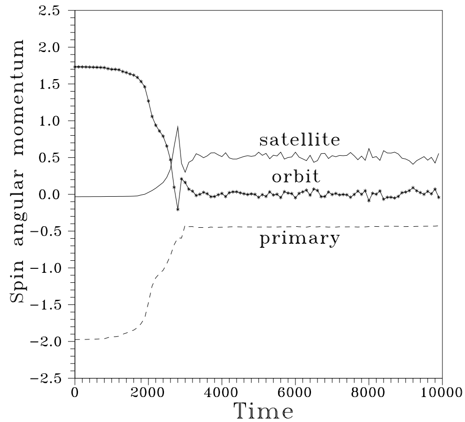



**Fig. 11.** The spin angular momentum of the primary galaxy (dashed line), of the satellite (solid line) and the orbital angular momentum (line with stars) as a function of time for ex. A.

galaxies)is plotted as a function of time for ex. A. The dashed line corresponds to the spin angular momentum of the primary galaxy, the solid line to the spin of the satellite and the solid line with stars to the angular momentum of the orbit with respect to the common center of mass. The spins along the other two directions are very close to zero. From this figure it is obvious that the orbit of the satellite is retrograde with respect to the rotation of the primary galaxy and that angular momentum is transferred from the orbit to the individual particles of the two galaxies. The spin of the primary galaxy starts from a negative value and is reduced by a factor of 4, until the remnant of the merging has relaxed (at a time of about 3000 $t_{un}$) and then remains constant. The satellite, on the other hand, starts with an almost zero initial value of spin angular momentum and after the relaxation acquires a positive spin angular momentum with approximately the same mean value with the value of the primary. These results are in agreement with Barcells and Quinn [2], who had mentioned that the survival of the satellite does not imply the survival of its spin. The spin of the satellite becomes always aligned with the orbital angular momentum, after the merger, irrespective of its initial orientation or sign. Therefore, the major effect for the presence of counterrotation is due to the orbital angular momentum of the satellite. Another parameter that could be useful to check and can be thought as a trace of a past merger event, is the ellipticity profile of the galaxy. Bender [4] pointed out that in some observed counterrotating galaxies, the radial transition between the core and the main body is clearly visible in their ellipticity profiles. He proposed that this should be a sign of a past merger event. However, the ellipticity profiles derived from our simulations (Fig. 12), both from merger and from cosmological initial conditions, look very much alike with the ones presented in Bender's paper. In Fig. 12 the line with dots correspond to the ellipticity of the merger remnant of ex. A as a function of it's longest axis, while the line with stars corresponds to the ellipticity of the galaxy with cosmological initial conditions.

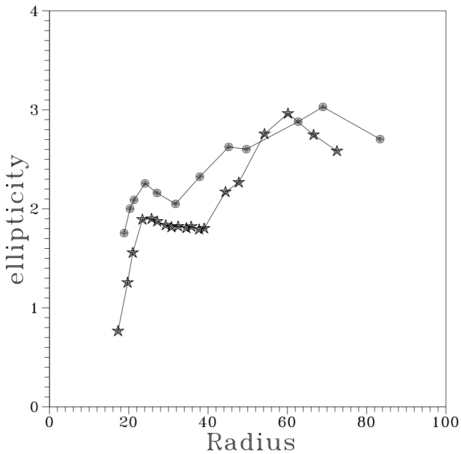



**Fig. 12.** The ellipticity as a function of the radius along the major axis of the remnant galaxy of ex. A.(line with dots) and for the galaxy formed by cosmological initial conditions (line with stars)

A peak of the radial profile of the ellipticity appears in both experiments and therefore one cannot use the ellipticity profile of an observed counterrotating galaxy to decide whether it is the product of a merger or it is produced from a single primordial collapse.

A theoretical argument below gives more chances to the direct scenario. In the case of cosmological initial conditions there are two main parameters that have to be arranged in order to produce counterrotating galaxies, namely, the axial ratio of the bar-like density excess, and the strength of the initial tidal field. Statistical estimations of these parameters can be directly derived from the power spectrum of the initial density perturbations (and its various moments).

On the other hand, only a small number of merger events can guarantee the formation of counterrotating remnants. Those in which a considerably large number of parameters are well tuned.

For example:

a) The mass ratio of the satellite and the primary as well as the density profile of the satellite must be inside a certain range of values, so that the satellite not only survives into the center but also dominates the core of the remnant.

b) the satellite's orbit must be on the rotation plane of the primary, otherwise the dynamical friction between the satellite and the rotating primary causes the satellite to pivot such as to be oriented almost upright in relation to the rotation plane of the primary,

c) the orbit of the satellite must be retrograde with respect to the rotation of the primary and have initial kinetic energy such that it does not escape, disrupted, or totally reverse the sign of the rotation throughout the radius of the remnant.

These arguments considerably reduce the possibilities of successful merger events. A question therefore is how to assign the formation of so many observed counterrotating ellipticals in such a demanding mechanism.

Recently a study has been made for the environment of galaxies in which there is gas or stellar counterrotation by Bettoni et al [9]. They concluded that no significant differences appear between the environments of counterrotating and normal galaxies. Therefore the hypothesis that counterrotating galaxies and polar rings derived from a recent interaction with a small satellite or a galaxy of similar size seems to be disproved and these galaxies seem to follow the idea that all galaxies are born through a merger process of smaller objects occurring very early in their life (cosmological initial conditions), or that they have been derived from a continuous infall of gas that formed stars later.

**Acknowledgements**

This work was supported by the Research Committee of the Academy of Athens (grant 200/517). We wish to thank Dr. Sv. Aarseth for his code.

## References

1. Bak, J.: Thesis: “Retrograde minor mergers and counterrotating cores” (2000)
2. Barcells, M., and Quinn, P.J.: ApJ **361**, 381 (1990)
3. Barcells, M. and Gonzalez, A.C.: ApJ **505**, L109 (1998)
4. Bender R.: A&A **202**, L5 (1988)
5. Bender, R., and Surma P.: A&A **258**, 250 (1992)
6. Bertola, F. and Bettoni, D.: ApJ **329**, 102 (1988)
7. Bertola, F., Buson, L.M., and Zeilinger, W.W.: Nature, **335**, 705 (1988)
8. Bettoni, D., Fasano, G., and Galletta, G.: Astron. J. **99**, 1789 (1990)
9. Bettoni, D., Galleta, G., and Prada F.: A&A **374**, 83 (2001)
10. Chandrasekhar, S.: ApJ **97**, 255 (1943)
11. Ciri, R., Bettoni, D., and Galletta, G.: Nature **375**, 661 (1995)
12. Duncan, M., Farouki R., and Shapiro S.: ApJ **271**, 22 (1983)
13. Efstathiou, G., Ellis R.S. and Carter D.: MNRAS **201**, 975 (1982)
14. Franx, M., and Illingworth G.: ApJ **327**, L55 (1988)
15. Franx, M., Illingworth G. and Heckman T.: ApJ **344**, 613 (1989)
16. Harsoula, M.: Thesis: “Cosmological collapses and formation of counterrotating galaxies” (1999)
17. Harsoula, M. and Voglis, N.: A&A **335**, 431 (1998)
18. Hau, G.K.T. and Thomson R.C.: MNRAS **270**, L23 (1994)
19. Hausman, M., and Ostriker J.: ApJ **224**, 320 (1978)
26. Hernquist, L., Barnes, J.: Nature **354**, 210 (1991)
21. Kormendy, J.: ApJ **287**, 577 (1984)
22. Kuijken, K., Fisher D., Merrifield M.R.: MNRAS, **283**, 543 (1996)
23. Mehlert, D., Saglia R.P., Bender R. and Wegner G.: A&A **332**, 33 (1998)
24. Negroponte, J., White, S.: MNRAS **205**, 1009 (1983)
25. Peebles P.J.E.: ApJ **155**, 393 (1969)
26. Rubin, V.C., Graham, J.A., and Kenney J.D.P.: ApJ **394**, L9 (1992)
27. Voglis, N., Hiotelis, N.: A&A **218**, 1 (1989)

28. Voglis, N., Hiotelis, N., Hoeflich, P.: A&A **249**, 5 (1991)
29. Voglis., N., Harsoula, M., Efthymiopoulos, Ch.: Cel. Mech. Dyn. Astr. **78**, 265-278 (2000)
30. Voglis, N. and Harsoula, M.: in *Modern Theoretical and Observational Cosmology*, Plionis M. and Kotsakis S. (eds), Kluwer Academic Publishers, 276 (2001).
31. Zel'dovich, Ya.: A&A **5**, 84 (1970)

# Pitch Angle of Spiral Galaxies as Viewed from Global Instabilities of Flat Stellar Disks

Shunsuke Hozumi

Max-Planck-Institut für Astronomie, Königstuhl 17, D-69117 Heidelberg, Germany

**Abstract.** We investigate whether the behavior of the pitch angle, predicted by a local dispersion relation derived from the density wave theory, can be applied to that of the fastest growing, globally unstable modes of flat stellar disks. We pay attention to two-armed modes, and obtain such global modes by numerically integrating the linearized collisionless Boltzmann equation. The results show that the pitch angle of the fastest growing modes has a tight correlation with the ratio of the square of the radial velocity dispersion to the surface density, as indicated by the local theory, over a sufficiently wide range of radii. The correspondence between the global modes and global properties of spiral galaxies is briefly discussed.

## 1 Introduction

Spiral structure is one of the most prominent features in disk galaxies. In fact, Hubble [6] classified spiral galaxies on the basis of the observed appearance of spiral arms. However, what his classification scheme, known as the Hubble sequence, represents physically remains still unclear.

One of the most successful theories to understand spiral structure is the density wave theory, which was originally conceived by Lindblad [10], and was first formulated by Lin and Shu [9]. This theory was based on local analysis, so that owning to the inhomogeneous density distributions in galaxy disks, density waves are propagated with a group velocity, and disappear in a few dynamical times [20]. In addition, it is found that spiral shocks [15] induced by a gaseous component in the disk damp the underlying spiral potential, again on a dynamical time scale [8].

The difficulties confronted with the density wave theory mentioned above may be overcome by the concept of global modes. Indeed, global analysis of thin disks has revealed that many unstable modes exist in self-gravitating disks treated with a fluid approximation [1] and in those composed of stars [7]. In particular, Bertin et al. [2] have shown that all Hubble morphological types can be realized by global modes using fluid disk models with appropriate basic states. In their modal calculations, it is demonstrated that the pitch angle of global modes depends on a fraction of the active disk mass, the distribution of Toomre's $Q$ parameter [19], the typical $Q$ value, and so on. Since their major concern was to specify the basic state to generate and support a given spiral structure, they did not describe explicitly what determined the pitch angle of spiral arms.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 380–386, 2003.

Since the density wave theory gives, in general, a good description of real spiral galaxies [17], it can help us understand the physics of spiral structure in spite of a local theory based on a WKBJ approximation. However, it is not necessarily clear whether the properties of global modes can be fully described by local density waves (see [21]). In this paper, we first show what is predicted for the pitch angle of unstable modes on the basis of the density wave theory. Next, this prediction is examined for the fastest growing two-armed modes of stellar disks in a global context.

## 2 Prediction by the Density Wave Theory

For tightly wrapped spiral waves in a cool disk, the dispersion relation for a one-component fluid model [3] is given by

$$(\omega - m\Omega)^2 = c^2 k^2 - 2\pi G \Sigma |k| + \kappa^2, \tag{1}$$

where $\omega$, $\Omega$, $\kappa$, $k$, $c$, $G$, $\Sigma$, and $m$ are the wave frequency, angular speed of a star, epicyclic frequency, wavenumber, sound speed, gravitational constant, surface density of the disk, and number of arms in spiral patterns, respectively. Equation (1) is rewritten as

$$(\omega - m\Omega)^2 = c^2(|k| - \pi G\Sigma/c^2)^2 + \kappa^2 \left(1 - 1/Q_{\mathrm{g}}^2\right), \tag{2}$$

where $Q_{\mathrm{g}} = \kappa c/(\pi G \Sigma)$ is the Toomre stability parameter for gas disks. When $Q_{\mathrm{g}} > 1$, all waves are stable, so that there is no specific wave to be selected. However, taking into account unstable waves, the wave mode with the highest growth rate should appear in the disk by overwhelming the others with time. Since the Toomre parameter is a criterion valid only for local axisymmetric Jeans instabilities, we neglect the stability condition that $Q_{\mathrm{g}} > 1$ in the case of non-axisymmetric instabilities like those two-armed modes focused on here, and will find what is expected if (2) includes unstable waves.

Equation (2) indicates that the most unstable wave has the wavenumber $|k_{\mathrm{m}}| = \pi G\Sigma/c^2$, which is converted to the wavelength $\lambda_{\mathrm{m}}$ given by

$$\lambda_{\mathrm{m}} = 2c^2/G\Sigma. \tag{3}$$

Thus, we can see that the wavelength of the most unstable mode depends on $c^2/\Sigma$. As the wavelength is larger, the pitch angle of spiral arms becomes larger. Therefore, the density wave theory predicts, as is well-known, that the pitch angle increases with increasing $c^2/\Sigma$.

## 3 Models and Method

Since we aim not at the reproduction of real spiral structures but at demonstrating how well and to what extent the local theory gives a representation of the features of global modes, we use disk models appropriate for our concern, irrespective of the deviation from realism.

We adopt infinitesimally thin Toomre disks [18] without bulges, whose surface density distributions, $\Sigma$, and potentials, $\Psi$, are given, respectively, by

$$\Sigma(R) = \frac{Mq}{2\pi a^2}\left(1 + \frac{R^2}{a^2}\right)^{-3/2}, \tag{4}$$

and

$$\Psi(R) = -\frac{GM}{\sqrt{R^2 + a^2}}, \tag{5}$$

where $R$ is the distance from the center of the disk, and $a$ is the length scale. Here, $q$ $(0 < q \leq 1)$ represents an active disk mass fraction of the total mass. Since the Toomre disks are the flat version of Plummer's models [14] in three-dimensional configurations, for each disk model the fraction, $1 - q$, of the total mass is considered to reside in a surrounding dark halo represented by a Plummer model with the same length scale $a$ as that of the disk. Thus, in our models, the total mass distribution including a disk and a halo is identical, and the mass fraction of the disk is different.

For this mass profile, we can use Miyamoto's distribution functions (DFs) [12] with respect to directly rotating stars, $F^+(\varepsilon, j)$, where $\varepsilon$ and $j$ are the energy and angular momentum of a star per unit mass, respectively. Retrograde stars are introduced in the same manner as that adopted by Nishida et al. [13]. Then, the equilibrium DFs, $F_0$, are given by

$$F_0(\varepsilon, j) = \begin{cases} q[(1/2)F_0^+(\varepsilon) + F_1^+(\varepsilon, j)], & (j \geq 0) \\ (1/2)qF_0^+(\varepsilon), & (j < 0), \end{cases} \tag{6}$$

where the functions, $F_0^+(\varepsilon)$ and $F_1^+(\varepsilon, j)$, are derived from the expansion of $F^+(\varepsilon, j)$ as

$$F^+(\varepsilon, j) = F_0^+(\varepsilon) + F_1^+(\varepsilon, j). \tag{7}$$

The Miyamoto DFs [12] are characterized by the parameter $n$ that specifies the distribution of radial velocity dispersion, $c_{\mathrm{r}}$, such that

$$c_{\mathrm{r}}^2(R) = -\frac{1}{2n + 4}\Psi(R). \tag{8}$$

This equation indicates that $c_{\mathrm{r}}$ can be provided independently of the disk surface density.

If disk models satisfy the condition such that $(n + 2)q =$ constant, they have the same $c_{\mathrm{r}}^2/\Sigma$ distribution throughout the disk, as derived from (4), (5), and (8). We take $(n + 2)q = 5$ with the pairs of $(n, q) = (3, 1)$, $(4, 5/6)$, $(5, 5/7)$, $(6, 5/8)$, $(7, 5/9)$, and $(10, 5/12)$, and $(n + 2)q = 3$ with the pairs of $(n, q) = (3, 3/5)$, $(4, 1/2)$, $(5, 3/7)$, and $(6, 3/8)$. In Fig. 1, the distributions of the $Q$ parameter for both model sequences are presented. Here, $Q$ is defined for stellar disks as

$$Q = \frac{\kappa c_{\mathrm{r}}^2}{3.36 G\Sigma}. \tag{9}$$

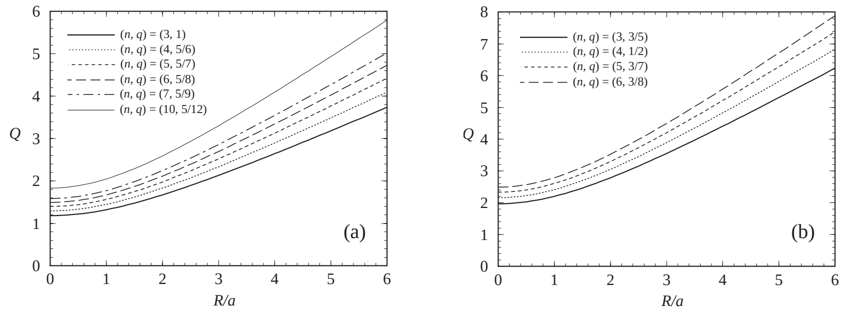



**Fig. 1.** Distribution of the Toomre stability parameter $Q$ for (a) the models with $(n+2)q = 5$, and for (b) the models with $(n+2)q = 3$

It is found from this figure that all models are locally stable, i.e., $Q > 1$.

The most unstable global modes are obtained by numerically integrating the linearized collisionless Boltzmann equation given by

$$\frac{df_m}{dt} = \frac{\partial \psi_m}{\partial R}\frac{\partial F_0}{\partial u} + im\psi_m \frac{\partial F_0}{\partial j}, \tag{10}$$

where $f_m$ and $\psi_m$ are the perturbed distribution function and perturbed potential, respectively, with respect to the $m$-armed mode, and $i$ is the unit of the imaginary part. Equation (10) is solved as an initial value problem by evolving an arbitrary form of perturbation, imposed initially, forward in time until it has reached the exponential growth in $f_m$ and $\psi_m$. We restrict ourselves only to $m = 2$ modes, which are usually the most unstable modes among various $m$-armed modes. In fact, most of grand-design spirals show two-armed features. The numerical details are described in [5].

The units of mass and length, and the gravitational constant are taken so that $M = a = 1$ and $G = 1$, respectively. Then, the unit of time is $(a^3/GM)^{1/2}$.

## 4 Results

We have obtained the fastest growing two-armed modes of the models with $(n+2)q = 5$ and those with $(n+2)q = 3$. The growth rate, pattern speed, and corotation radius of each model are summarized in Table 1 for $(n+2)q = 5$, and in Table 2 for $(n+2)q = 3$.

We calculate the pitch angle, $i_{\mathrm{p}}$, given by

$$\cot i_{\mathrm{p}} = \left| mR\frac{\partial \phi}{\partial R} \right|, \tag{11}$$

where for a specified arm in the disk, $\phi$ is the phase angle of the density crest of the mode along $R$ [3]. Here, we employ $m = 2$.

The pitch angle of each mode, measured in degrees, is shown in Fig. 2, which illustrates that the pitch angle profile does not change substantially from model

**Table 1.** Results for the fastest growing modes of the models with $(n+2)q = 5$

| $n$ | $q$ | Growth Rate | Pattern Speed | Corotation Radius |
|---|---|---|---|---|
| 3 | 1 | 0.224 | 0.333 | 1.83 |
| 4 | 5/6 | 0.167 | 0.328 | 1.85 |
| 5 | 5/7 | 0.119 | 0.313 | 1.92 |
| 6 | 5/8 | 0.0851 | 0.294 | 2.03 |
| 7 | 5/9 | 0.0652 | 0.275 | 2.15 |
| 10 | 5/12 | 0.0516 | 0.228 | 2.49 |

**Table 2.** Results for the fastest growing modes of the models with $(n+2)q = 3$

| $n$ | $q$ | Growth Rate | Pattern Speed | Corotation Radius |
|---|---|---|---|---|
| 3 | 3/5 | 0.0309 | 0.226 | 2.51 |
| 4 | 1/2 | 0.0158 | 0.213 | 2.62 |
| 5 | 3/7 | 0.00744 | 0.201 | 2.74 |
| 6 | 3/8 | 0.00449 | 0.190 | 2.86 |

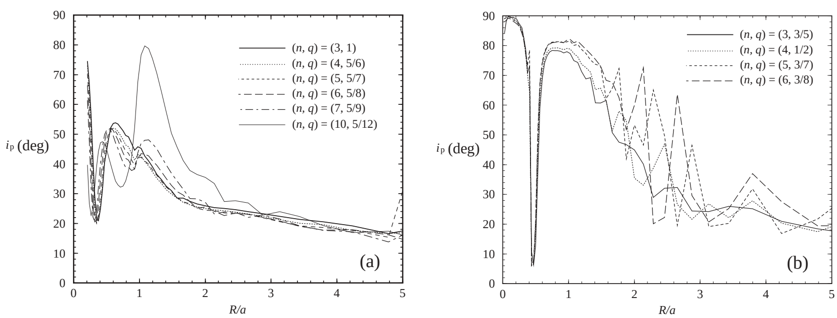



**Fig. 2.** Pitch angle of the fastest growing two-armed modes for (a) the models with $(n+2)q = 5$, and for (b) the models with $(n+2)q = 3$

to model for each model sequence. For the $(n+2)q = 5$ sequence, all but the model with $(n, q) = (10, 5/12)$ have nearly the same pitch angle profile within $R = 5$. Even if the model with $(n, q) = (10, 5/12)$ is included in this sequence, the difference in pitch angle is rather small from $R \sim 2$ to $R \sim 5$. For the $(n+2)q = 3$ sequence, the pitch angles are not well determined because the very small growth rates do not enhance the density contrast of the unstable modes conspicuously. In spite of the large scatters, the pitch angles of these models appear to fluctuate around a certain profile. Therefore, also in this sequence, it can be found that the pitch angle profiles are approximately similar to one another. These results

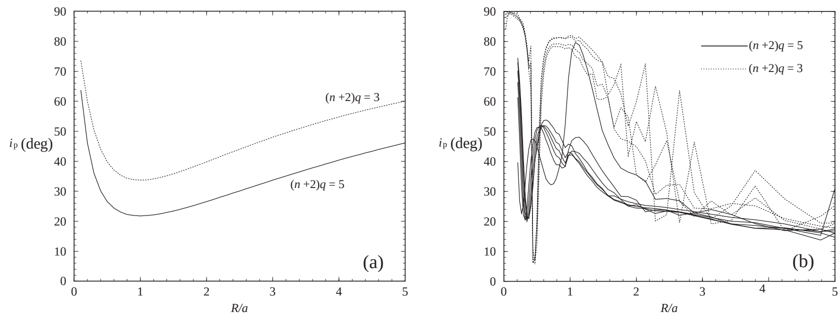



**Fig. 3.** Pitch angle of unstable two-armed modes for (a) expected profiles calculated from the local dispersion relation for $(n+2)q = 5$ (solid line), and for $(n+2)q = 3$ (dotted line), and for (b) obtained profiles of all the models with $(n+2)q = 5$ (solid lines) and $(n+2)q = 3$ (dotted lines)

are remarkable in that models with different radial velocity dispersion profiles and different degrees of self-gravity show almost the same pitch angle profile over a sufficiently wide range of radii.

According to the local theory prediction, the increase in $c_{\mathrm{r}}^2/\Sigma$ leads to the increase in pitch angle, as found from (3) and (11). In our models, $c_{\mathrm{r}}^2/\Sigma$ is inversely proportional to $(n+2)q$, and so, the pitch angles for the $(n+2)q = 3$ sequence should be systematically larger than those for the $(n+2)q = 5$ sequence, as indicated by Fig. 3 (a), in which the pitch angles are calculated from (3) and (11). We can see this expected behavior in Fig. 3 (b), where the pitch angle profiles of all the models are put together. However, the difference in pitch angle for both model sequences becomes very small at larger radii than $R = 3$ unlike those values obtained from the local dispersion relation which are presented in Fig. 3 (a).

## 5 Discussion and Conclusions

We have found that the pitch angles of the fastest growing modes obey the prediction derived from the density wave theory even in a global context. We conclude from our global mode calculations that the pitch angles of spiral arms are regulated to a considerable degree by a typical value of $c_{\mathrm{r}}^2/\Sigma$, if spiral structure is viewed as a manifestation of the fastest growing modes of stellar disks. In addition, our results indicate that the density wave theory is helpful to understand the properties of global modes qualitatively. In particular, although the theory is developed for tightly wound spirals, it is still useful for loosely wound spirals because the pitch angles obtained here are typically 20°.

Roberts and Haynes [16] have summarized the global properties of spiral galaxies along the Hubble sequence. According to their Fig. 2, there is no practically systematic change in total mass at least from Sa to Sc. Furthermore, their Fig. 3 illustrates that the total surface density including gaseous and stellar components decreases slightly from Sa to Sc. If the mass distributions of spiral galaxies would not change substantially, the decrease in total surface density and the increase in pitch angle, together with our results, suggest that $c_{\mathrm{r}}$ would be nearly constant or increase from Sa to Sc. Therefore, observations of $c_{\mathrm{r}}$ along the Hubble sequence will clarify the relevance of our argument.

Recently, Ma [11] has revealed that the pitch angle increases as the disk surface density decreases by analyzing all Hubble type galaxies altogether. Ma's finding is consistent with our results of global mode calculations. His analysis shows a large scatter in the relation between the pitch angle and the surface density. If a typical $c_{\mathrm{r}}^2/\Sigma$ instead of $\Sigma$ was chosen for each spiral galaxy, this scatter might be reduced.

From the arguments mentioned above, real spiral structures appear to reflect the characteristics of the fastest growing, global modes of stellar disks. In fact, the old stellar population observed at near-infrared wavelengths forms large-scale smooth symmetric arms in the disk (e.g., [4]). These features may be a manifestation of globally unstable modes.

## References


1. S. Aoki, M. Noguchi, M. Iye: Publ. Astron. Soc. Jpn. **31**, 737 (1979)
2. G. Bertin, C. C. Lin, S. A. Lowe, R. P. Thurstans: Astrophys. J. **338**, 78 (1989)
3. J. Binney, S. Tremaine: *Galactic Dynamics* (Princeton University Press, Princeton 1987) Chap. 6
4. D. L. Block, G. Bertin, A. Stockton, P. Grosbøl, A. F. M. Moorwood, R. F. Peletier: Astron. Astrophys. **288**, 365 (1994)
5. S. Hozumi, T. Fujiwara, M. T. Nishida: Publ. Astron. Soc. Jpn. **39**, 447 (1987)
6. E. P. Hubble: Astrophys. J. **64**, 321 (1926)
7. A. J. Kalnajs: Astrophys. Lett. **11**, 41 (1972)
8. A. J. Kalnajs: Astrophys. J. **175**, 63 (1972)
9. C. C. Lin, F. H. Shu: Astrophys. J. **140**, 646 (1964)
10. B. Lindblad: Stockholm Observ. Ann. **22**, 3 (1963)
11. J. Ma: Astron. Astrophys. **388**, 389 (2002)
12. M. Miyamoto: Publ. Astron. Soc. Jpn. **23**, 21 (1971)
13. M. T. Nishida, Y. Watanabe, T. Fujiwara, S. Kato: Publ. Astron. Soc. Jpn. **36**, 27 (1984)
14. H. C. Plummer: Mon. Not. R. Astro. Soc. **71**, 460 (1911)
15. W. W. Roberts: Astrophys. J. **158**, 123 (1969)
16. M. S. Roberts, M. P. Haynes: Annu. Rev. Astron. Astrophys. **32**, 115 (1994)
17. W. W. Roberts, Jr., M. S. Roberts, F. H. Shu: Astrophys. J. **196**, 381 (1975)
18. A. Toomre: Astrophys. J. **138**, 385 (1963)
19. A. Toomre: Astrophys. J. **139**, 1217 (1964)
20. A. Toomre: Astrophys. J. **158**, 899 (1969)
21. A. Toomre: 'What amplifies the spirals'. In: *The Structure and Evolution of Normal Galaxies*, ed. by S.M. Fall, D. Lynden-Bell (Cambridge University Press, Cambridge 1981) pp. 111–136

# Collisionless Evaporation from Cluster Elliptical Galaxies

Veruska Muccione[1] and Luca Ciotti[2]

[1] Geneva Observatory, 51. ch. des Maillettes, 1290 Sauverny, Switzerland
[2] Dipartimento di Astronomia, Università di Bologna, via Ranzani 1, 40127 Bologna, Italy

**Abstract.** We describe a particular aspect of the effects of the parent cluster tidal field (CTF) on stellar orbits inside cluster Elliptical galaxies (Es). In particular we discuss, with the aid of a simple numerical model, the possibility that *collisionless stellar evaporation* from elliptical galaxies is an effective mechanism for the production of the recently discovered intracluster stellar populations (ISP). A preliminary investigation, based on very idealized galaxy density profiles (Ferrers density distributions), showed that over an Hubble time, the amount of stars lost by a representative galaxy may sum up to the 10% of the initial galaxy mass, a fraction in interesting agreement with observational data. The effectiveness of this mechanism is due to the fact that the galaxy oscillation periods near equilibrium configurations in the CTF are comparable to stellar orbital times in the external galaxy regions. Here we extend our previous study to more realistic galaxy density profiles, in particular by adopting a triaxial Hernquist model.

## 1 Introduction

Observational evidences of an Intracluster Stellar Population (ISP) are mainly based on the identification of *intergalactic* planetary nebulae and red giant stars (see, e.g., [1],[2],[3],[4],[5]). Overall, the data suggest that approximately 10% (or even more) of the stellar mass of clusters is contributed by the ISP [6]. The usual scenario assumed to explain the finding above is that gravitational interactions between cluster galaxies, and interactions between the galaxies with the gravitational field of the cluster, lead to a substantial stripping of stars from the galaxies themselves.

Here, supported by a curious coincidence, namely by the fact that *the characteristic times of oscillation of a galaxy around its equilibrium position in the cluster tidal field (CTF) are of the same order of magnitude of the stellar orbital periods in the external part of the galaxy itself*, we explore the effects of interaction between stellar orbits inside the galaxies and the CTF. In fact, based on the observational evidence that the major axis of cluster Es seems to be preferentially oriented toward the cluster center, N-body simulations showed that model galaxies tend to align, as observed, reacting to the CTF as rigid bodies [7] . By assuming this idealized scenario, a stability analysis then showed that this configuration is of stable equilibrium, and allowed to calculate the oscillation periods in the linearized regime [8]. In particular, oscillations around two stable equilibrium configurations have been considered, namely: 1) when the center

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 387–392, 2003.

of mass of the galaxy is at rest at center of a triaxial cluster, and the galaxy inertia ellipsoid is aligned with the CTF principal directions, and 2) when the galaxy center of mass is placed on a circular orbit in a spherical cluster, and the galaxy major axis points toward the galaxy center while the galaxy minor axis is perpendicular to the orbital plane.

Here, prompted by these observational and theoretical considerations, we extend a very preliminary study of the problem [9], by evolving stellar orbits in a more realistic galaxy density profile: for simplicity we restrict to case 1) above, while the full exploration of the parameter space, together with a complete discussion of case 2), will be given elsewhere [10]. It is clear, however, that both cases are rather exceptional. Most cluster galaxies neither rest in the cluster center nor move on circular orbits, but they move on elongated orbits with very different pericentric and apocentric distances from the cluster's center; in a triaxial cluster many orbits are boxes and some orbits can be chaotic. These latter cases can be properly investigated only by direct numerical simulation of the stellar motions inside the galaxies, coupled with the numerical integration of the equations of the motion of the galaxies themselves.

## 2 The Physical Background

Without loss of generality we assume that in the (inertial) Cartesian coordinate system $C$, with the origin on the cluster center, the CTF tensor $\boldsymbol{T}$ is in diagonal form, with components $T_i$ $(i = 1, 2, 3)$. By using three successive, counterclockwise rotations ($\varphi$ around $x$ axis, $\vartheta$ around $y'$ axis and $\psi$ around $z''$ axis), the linearized equations of the motion for the galaxy near the equilibrium configuration can be written as

$$\ddot{\varphi} = \frac{\Delta T_{32}\Delta I_{32}}{I_1}\varphi, \qquad \ddot{\vartheta} = \frac{\Delta T_{31}\Delta I_{31}}{I_2}\vartheta, \qquad \ddot{\psi} = \frac{\Delta T_{21}\Delta I_{21}}{I_3}\psi, \tag{1}$$

where $\Delta T$ is the antisymmetric tensor of components $\Delta T_{ij} \equiv T_i - T_j$, and $I_i$ are the principal components of the galaxy inertia tensor. In addition, let us also assume that $T_1 \geq T_2 \geq T_3$ and $I_1 \leq I_2 \leq I_3$, i.e., that $\Delta T_{32}, \Delta T_{31}$ and $\Delta T_{21}$ are all less or equal to zero (see, e.g., [8], [10]). Thus, the equilibrium position associated with (1) is *linearly stable*, and its solution is

$$\varphi = \varphi_{\mathrm{M}} \cos(\omega_\varphi t), \quad \vartheta = \vartheta_{\mathrm{M}} \cos(\omega_\vartheta t), \quad \psi = \psi_{\mathrm{M}} \cos(\omega_\psi t), \tag{2}$$

where

$$\omega_\varphi = \sqrt{\frac{\Delta T_{23}\Delta I_{32}}{I_1}}, \quad \omega_\vartheta = \sqrt{\frac{\Delta T_{13}\Delta I_{31}}{I_2}}, \quad \omega_\psi = \sqrt{\frac{\Delta T_{12}\Delta I_{21}}{I_3}}. \tag{3}$$

For computational reasons the best reference system in which calculate stellar orbits is the (non inertial) reference system $C'$ in which the galaxy is at rest, and its inertia tensor is in diagonal form. The equation of the motion for a star in $C'$ is

$$\ddot{\boldsymbol{x}}' = \mathcal{R}^{\mathrm{T}}\ddot{\boldsymbol{x}} - 2\boldsymbol{\Omega} \wedge \boldsymbol{v}' - \dot{\boldsymbol{\Omega}} \wedge \boldsymbol{x}' - \boldsymbol{\Omega} \wedge (\boldsymbol{\Omega} \wedge \boldsymbol{x}'), \tag{4}$$

where $\boldsymbol{x} = \mathcal{R}(\varphi, \vartheta, \psi)\boldsymbol{x}'$, and

$$\boldsymbol{\Omega} = (\dot\varphi \cos\vartheta \cos\psi + \dot\vartheta \sin\psi, -\dot\varphi \cos\vartheta \sin\psi + \dot\vartheta \cos\psi, \dot\varphi \sin\vartheta + \dot\psi). \tag{5}$$

In (4)

$$\mathcal{R}^{\mathrm{T}}\ddot{\boldsymbol{x}} = -\nabla_{\boldsymbol{x}'}\phi_{\mathrm{g}} + (\mathcal{R}^{\mathrm{T}}\boldsymbol{T}\mathcal{R})\boldsymbol{x}', \tag{6}$$

where $\phi_{\mathrm{g}}(\boldsymbol{x}')$ is the galactic gravitational potential, $\nabla_{\boldsymbol{x}'}$ is the gradient operator in $C'$, and we used the tidal approximation to obtain the star acceleration due to the cluster gravitational field.

## 3 Galaxy and Cluster Models

For simplicity we assume that the galaxy and cluster densities are stratified on homeoids. In particular, the galaxy density belongs to ellipsoidal generalization of the widely used $\gamma$-models ([11],[12]):

$$\rho_{\mathrm{g}}(m) = \frac{M_{\mathrm{g}}}{\alpha_1\alpha_2\alpha_3}\frac{3-\gamma}{4\pi}\frac{1}{m^{\gamma}(1+m)^{4-\gamma}}, \tag{7}$$

where $M_{\mathrm{g}}$ is the total mass of the galaxy, $0 \leq \gamma \leq 3$ and

$$m^2 = \sum_{i=1}^{3}\frac{(x_i')^2}{\alpha_i^2}, \qquad \alpha_1 \geq \alpha_2 \geq \alpha_3. \tag{8}$$

The inertia tensor components of a generic homeoidal density distribution (in the natural reference system adopted in (8)), are given by

$$I_i = \frac{4\pi}{3}\alpha_1\alpha_2\alpha_3(\alpha_j^2 + \alpha_k^2)h_{\mathrm{g}}, \tag{9}$$

where $h_{\mathrm{g}} = \int_0^\infty \rho_{\mathrm{g}}(m)m^4 dm$, and so $I_1 \leq I_2 \leq I_3$. Note that, from (3) and (9) it results that the frequencies for homeoidal stratifications *do not depend on the specific density distribution assumed*, but only on the quantities $(\alpha_1, \alpha_2, \alpha_3)$. We also introduce the two ellipticities

$$\frac{\alpha_2}{\alpha_1} \equiv 1 - \epsilon, \qquad \frac{\alpha_3}{\alpha_1} \equiv 1 - \eta, \tag{10}$$

where $\epsilon \leq \eta \leq 0.7$.

A rough estimate of *characteristic stellar orbital times* inside $m$ is given by $P_{orb}(m) \simeq 4P_{\mathrm{d}yn}(m) = \sqrt{3\pi/G\overline{\rho_{\mathrm{g}}}(m)}$, where $\overline{\rho_{\mathrm{g}}}(m)$ is the mean galaxy density inside $m$. We thus obtain

$$P_{orb}(m) \simeq 9.35 \times 10^6 \sqrt{\frac{\alpha_{1,1}^3(1-\epsilon)(1-\eta)}{M_{\mathrm{g},11}}}m^{\gamma/2}(1+m)^{(3-\gamma)/2} \quad \text{yrs}, \tag{11}$$

where $M_{\mathrm{g},11}$ is the galaxy mass normalized to $10^{11}M_\odot$, $\alpha_{1,1}$ is the galaxy "core" major axis in kpc units (for the spherically symmetric $\gamma = 1$ Hernquist model [13],

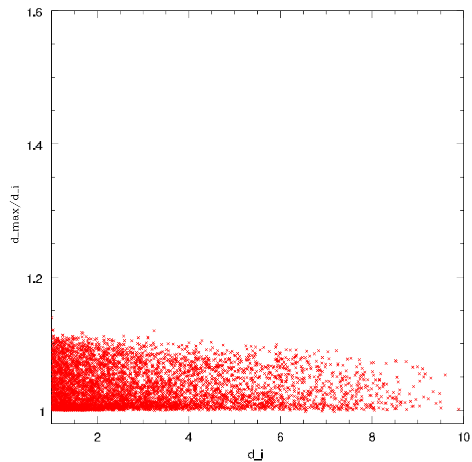



**Fig. 1.** Distribution of the $d_{\max}/d_{\rm i}$ ratio vs. $d_{\rm i}/\alpha_1$ after an Hubble time for the model galaxy at rest. $d_{\rm i}$ is the initial distance of the star from the galaxy center, while $d_{\max}$ is the maximum distance from the galaxy center reached during the simulation.

$R_{\rm e} \simeq 1.8\alpha_1$); thus, in the outskirts of normal galaxies orbital times well exceed $10^8$ or even $10^9$ yrs. For the cluster density profile we assume

$$\rho_{\rm c}(m) = \frac{\rho_{\rm c,0}}{(1+m^2)^2}, \tag{12}$$

where $m$ is given by an identity similar to (8), with $a_1 \geq a_2 \geq a_3$, and, in analogy with (10) we define $a_2/a_1 \equiv 1-\mu$ and $a_3/a_1 \equiv 1-\nu$, with $\mu \leq \nu \leq 1$. It can be shown (see, e.g., [8],[10]) that the CTF components at the center of a non-singular homeoidal distribution are given by

$$T_i = -2\pi G \rho_{\rm c,0} w_i(\mu,\nu), \tag{13}$$

where the dimensionless quantities $w_i$ are independent of the specific density profile, $w_1 \leq w_2 \leq w_3$ for $a_1 \geq a_2 \geq a_3$, and so the conditions for stable equilibrium in (1) are fulfilled ([8],[10]). The quantity $\rho_{\rm c,0}$ is not a well measured quantity in real clusters, and for its determination we use the virial theorem, $M_{\rm c}\sigma_{\rm V}^2 = -U$, where $\sigma_{\rm V}^2$ is the virial velocity dispersion, that we assume to be estimated by the observed velocity dispersion of galaxies in the cluster. Thus, we can now compare the galactic oscillation periods:

$$P_\varphi = \frac{2\pi}{\omega_\varphi} \simeq \frac{8.58\times 10^8}{\sqrt{(\nu-\mu)(\eta-\varepsilon)}}\frac{a_{1,250}}{\sigma_{\rm V,1000}}\ \text{yrs}\ ,$$

$$P_\vartheta = \frac{2\pi}{\omega_\vartheta} \simeq \frac{8.58\times 10^8}{\sqrt{\nu\eta}}\frac{a_{1,250}}{\sigma_{\rm V,1000}}\ \text{yrs}\ ,$$

$$P_\psi = \frac{2\pi}{\omega_\psi} \simeq \frac{8.58\times 10^8}{\sqrt{\mu\varepsilon}}\frac{a_{1,250}}{\sigma_{\rm V,1000}}\ \text{yrs}\ .$$

(for small galaxy and cluster flattenings, where $a_{1,250} = a_1/250$ kpc and $\sigma_{V,1000} = \sigma_V/10^3$ km/s, [10]) with the characteristic orbital times in galaxies. Thus, from (11) and (14abc), it follows that *in the outer halo of giant Es, stellar orbital times can be of the same order of magnitude as the oscillatory periods of the galaxies themselves near their equilibrium position in the CTF*. For example, in a relatively small galaxy of $M_{g,11} = 0.1$ and $\alpha_{1,1} = 1$, $P_{orb} \simeq 1$ Gyr at $m \simeq 10$ (i.e., at $\simeq 5R_e$), while for a galaxy with $M_{g,11} = 1$ and $\alpha_{1,1} = 3$ the same orbital time characterizes $m \simeq 7$ (i.e., $\simeq 3.5R_e$).

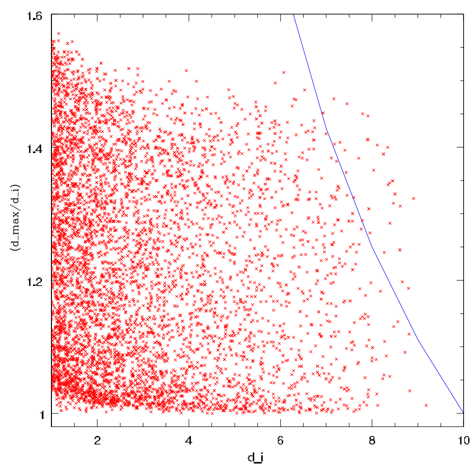



**Fig. 2.** Distribution of the $d_{\max}/d_i$ ratio vs. $d_i/\alpha_1$ after an Hubble time for the same galaxy model as in Fig. 1, when oscillating around its equilibrium position in the CTF.

In order to understand the effects of the galaxy oscillations on the stellar orbits, we performed a set of Monte-Carlo simulations, in which we followed the evolution of $10^4$ - $10^5$ "1-body problems" over the Hubble time by integrating numerically (4). At variance with [9], where we used simple and easy-to-integrate Ferrers density profiles, here we study orbital evolution in a more realistic (but also more demanding from the numerical point of view) galaxy density profile, namely a triaxial Hernquist model, obtained by assuming $\gamma = 1$ in (7). The gravitational potential inside the galaxy density distribution, in a form suitable for the numerical integration, was obtained by using an expansion technique useful in case of small density flattenings ([10],[14]). The initial conditions are generated by using the Von Neumann rejection method in phase-space (for details see [10]): note that, at variance with the analysis [9], now "stars" are characterized by initial velocities that can be different from zero. The code, a double-precision fortran code based on a Runge-Kutta scheme, runs on GRAVITOR, the Geneva Observatory 132 processors Beowulf cluster (**http://obswww.unige.ch/˜pfennige/gravitor/gravitor_e.html**). The computation of $10^4$ orbits usually requires 2 hours when using 10 nodes.

## 4 Preliminary Results and Conclusions

We show here, as an illustrative case, the behavior of the ratio $d_{\max}/d_{\rm i}$ as a function of $d_{\rm i}/\alpha_1$, for a moderately flattened galaxy model ($\epsilon \simeq 0.2$ and $\eta \simeq 0.3$), with $M_{\rm g} = 10^{11}\, M_{\odot}$, semi-mayor axis $\alpha_1 = 3$ kpc, and maximum oscillation angles equals to 0.1 rad. The cluster parameters are $a_{1,250} = \sigma_{\rm V,1000} = 1$, $\mu = 0.2$, $\nu = 0.4$, and the total number of explored orbits is $N_{\rm tot} = 10^4$. In order to show the effect of oscillations, in the following simulations we artificially eliminated the *direct* contribution of the CTF, as given by the second term in the r.h.s. of (6).

In Fig. 1 we show the result of a first simulation in which the galaxy is *not* oscillating: obviously, the ratio $d_{\max}/d_{\rm i}$ is in general (slightly) larger than unity, due to the initial velocity of each star. In Fig. 2 we show the result for the same galaxy model, when oscillating around the equilibrium position: the effects of the galaxy oscillations are clearly visible as a global "expansion" of the galaxy. As a reference, the solid line indicates the expansion ratio required to reach the representative distance of $10R_{\rm e}$ from the galaxy center. Thus, it is clear that the galaxy oscillations are certainly able to substantially modify the galaxy density profile. In particular, it will be of interest the study of the (more realistic) case in which the galaxy is in rotation around the cluster center. In this case we expect a different behavior of stellar orbits as a function of the distance of the galaxy center of mass from the cluster center: in fact, while inside the cluster core the CTF is *compressive* (see, e.g., [7],[8]), outside the CTF is *expansive* along the cluster radial direction, and in this latter case its direct effect should increase the expansive effect due to the galaxy oscillations. These cases are discussed in detail in [10].

## References


1. T.Theuns, S.J. Warren: MNRAS, 284 (1996)
2. R.H. Méndez et al: ApJ, 491 (1998)
3. J.J. Feldmeier , R. Ciardullo, G.H. Jacoby: ApJ, 503 (1998)
4. M. Arnaboldi, J.A.L. Aguerri, N.R. Napolitano, O. Gerhard, K.C. Freeman, J. Feldmeier , M. Capaccioli, R.P. Kudritzki, R.H. Méndez: AJ, 123 (2002)
5. P.R. Durrell, R. Ciardullo, J.J. Feldmeier, G.H. Jacoby, S. Sigurdsson: ApJ, 570 (2002)
6. H.C. Ferguson, N.R. Tanvir, T. von Hippel: Nature, 391 (1998)
7. L. Ciotti, S.N. Dutta: MNRAS, 270 (1994)
8. L. Ciotti, G. Giampieri: Cel. Mech. & Dyn. Astr., 68 (1998)
9. V. Muccione, L. Ciotti: Mem. S.A.It., in press (2003)
10. L. Ciotti, V. Muccione: in preparation
11. W. Dehnen: MNRAS, 256 (1993)
12. S. Tremaine et al.: AJ, 107 (1994)
13. L. Hernquist: ApJ, 356, 359 (1990)
14. L. Ciotti, G. Bertin: in preparation

# Chaos in Solar System Dynamics

Rudolf Dvorak

Institute of Astronomy, University of Vienna, Türkenschanzstrasse 17, A-1180 Vienna, Austria

**Abstract.** The effect of solar system chaos on small bodies, such as comets and asteroids, is quite different from that experienced by the major planets. It is more obvious in the motion of the large number of small bodies in the planetary region. Thus, the existence of different groups of comets and asteroids is due to the different qualities of the various resonances – mean motion resonances, secular resonances and three-body resonances – but especially because of resonance overlap. Moreover, chaotic motion has also been found in the motion of the planets and appears to be present on even a larger level in extra-solar planetary systems.

## 1 Introduction

When dealing with the celestial mechanics of solar system bodies we have to be aware that there is a fundamental difference from galactic dynamics even though the governing physical laws – primarily gravitation and also in special cases relativity – are the same. We can directly observe the motion of solar system bodies on time scales as short as years, months and even days. Furthermore, there is one special planetary system we are dealing with although now we have knowledge of about 100 extrasolar planets. For these systems the observational conditions are very different from observations in our planetary environment. In galactic dynamics we cannot observe directly the motion, for centuries the picture does not change for the observer on the Earth; we thus have a "snapshot" of the dynamics of our galaxy. On the contrary we have observations of galaxies in different stages of their age and can therefore deduce information on the dynamics of our own Galaxy. Furthermore, although a galaxy consists of billions of stars, we can treat their motions in the corresponding galactic potential as a motion in a "simple" Hamiltonian using an averaged gravitational potential. In solar system dynamics we have to deal with an n-body system where n is small (depending on the number of planets we wish to take into account). As we will see in this text the model of an n-body system has the advantage that it can be treated as an integrable dynamical system (namely the Keplerian motion of any body around the Sun, for a satellite around the parent planet) perturbed by the other bodies with significantly smaller masses than the Sun.

There are very good recent reviews on this subject ([15], [17]) which point out quite well the fundamental rôle of chaos in the dynamics of the solar system. In the present text we will, besides repeating more or less the fundamental ideas given in these two papers, critically give light to some of the results presented

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 395–411, 2003.

therein and report on new results concerning the rapidly increasing number of dynamical studies of extrasolar planetary systems.

One of the fundamental questions of astrodynamics is whether the solar systems is stable or not. Perturbation theory cannot give a final answer because of the limited time scales over which such theories are valid. It is also clear that straightforward numerical integration of the equations of motion cannot give a definitive answer because the inevitable accumulation errors lead to a deviation of the computed orbits from the "real" ones. An additional point of weakness is the incomplete dynamical model, which cannot perfectly represent the "real" solar system, even when all known effects, such as the gravitational forces of the largest minor planets [1], the oblateness of the Sun and the planets, the galactic field and the relativistic effects are included. Additionally there exists an intrinsic effect in nonlinear dynamical systems which makes the solutions uncertain owing to extreme sensitivity with respect to the initial conditions, known to physicists as *deterministic chaos*. This was pointed out for galactic dynamics in the seminal paper of Hénon and Heiles [9] and was found later through the work of Wisdom in the chaotic motion of asteroids near the 3:1 resonance with Jupiter [25].

The following sections are devoted to basic considerations of the dynamics of an n-body system with a dominating central mass, the appearance of chaotic motion due to the nonlinearity of the equations of motion and the action of different kinds of resonances, the motion of comets and asteroids, the long-term evolution of the orbits of the planets and the special orbit of Mercury. The conclusion summarizes the rôle of chaotic motion in solar systems dynamics. In an epilog some interesting results for extrasolar planetary systems are presented and show how our own system may serve as a "toy model" for planetary systems.

## 2 The n-Body Solar System

In heliocentric coordinates the equations of motion for the planets (modelled as mass points $m_\mathrm{i}$) may be written as

$$\ddot{\boldsymbol{q}}_\mathrm{i} = k^2 \left( -\frac{m_1 + m_\mathrm{i}}{r_\mathrm{i}^3} \boldsymbol{q}_\mathrm{i} + \sum_{j=2, j \neq i}^{\mathrm{n}} m_\mathrm{j} \left( \frac{\boldsymbol{q}_\mathrm{j} - \boldsymbol{q}_\mathrm{i}}{r_\mathrm{ji}^3} - \frac{\boldsymbol{q}_\mathrm{j}}{r_\mathrm{j}^3} \right) \right) . \tag{1}$$

Here the first term describes two-body motion around the Sun, and the second term is a small perturbing acceleration due to the presence of the other planets (with masses $m_\mathrm{j}$ of the order of $10^{-3}$ to $10^{-7}$ of the Sun's dominating mass). The appearance of the $3^{\mathrm{rd}}$ power of the distances $\mid \boldsymbol{q}_\mathrm{j} - \boldsymbol{q}_\mathrm{i} \mid = r_\mathrm{ji}$ in the denominator may cause large accelerations, even though the masses of the planets are small, when these distances become small themselves. In planetary theories this is not a practical problem because planets move on well separated orbits. It is different

[1] which are in fact taken into account for computing short time ephemerides of the planets

for comets, which often come quite close to planets (especially Jupiter) which can change their orbits significantly. Consequently comets having initially parabolic orbits may end up on elliptic orbits with moderate eccentricities. Besides the step-by-step numerical integration method (e.g. with a classical Runge-Kutta or a Lie series method [8]) a special method of solving the equations of motion has been developed in the last centuries by mathematicians and astronomers (Lagrange, Laplace, Brown, etc.). In perturbation theory one works with complicated series expansions including perhaps thousands of terms. The solutions are such that substituting the time in the series (also in form of a Fourier series) immediately yields the position in space (and in the sky). This method will be discussed briefly, as it gives deep insights into the nature of the dynamics of the planetary system and reveals the fundamental rôle of resonances.

### 2.1 Outline of the Classical Perturbation Method

For an approximate solution of the orbit of a planet the theory of two-body-motion can be applied which leads to six constants known as the orbital elements of a planet $\boldsymbol{\sigma} = (a, e, i, \omega, \Omega, T)^{\mathrm{T}}$ . Under the attraction of the other planets these elements change slowly in time; because the other masses are orders of magnitudes smaller than the Sun these changes are small. One can thus describe the motion of the planet $i$ by a $1^{\mathrm{st}}$ order differential equation of the form

$$\frac{d\boldsymbol{\sigma}_{\mathrm{i}}}{dt} = \sum_{j=1, j\neq i}^{n} F_{\mathrm{ij}}(\boldsymbol{\sigma}_{\mathrm{i}}, \boldsymbol{\sigma}_{\mathrm{j}}, \frac{\partial R_{\mathrm{ij}}}{\partial \boldsymbol{\sigma}_{\mathrm{i}}}) \tag{2}$$

where $R_{\mathrm{ij}}$ is the so-called perturbing function, which can be written as a Fourier-series with respect to the time. For a single planet of mass $m_1$ perturbed by another planet with mass $m_2$ the perturbing function is the following

$$R_{12} = m_2 \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} C_{\mathrm{j_1,j_2}} \cos[(j_1 n_1 + j_2 n_2)t + D_{\mathrm{j_1,j_2}}]. \tag{3}$$

The perturbations of the other planets may just be added, leading to a perturbing function $R_1 = R_{12} + R_{13} + \ldots + R_{1\mathrm{n}}$ depending on the number of planets taken into account; this simplification holds only in a first order theory. In a rather comprehensive formulation the Delaunay elements can be used

$$\begin{aligned} L_1 &= \kappa_1 \sqrt{a} & l_1 &= M_1 = n_1 t \\ G_1 &= L_1 \sqrt{(1-e^2)} & g_1 &= \omega_1 \\ H_1 &= G_1 \cos i & h_1 &= \Omega_1 \end{aligned} \tag{4}$$

where $\kappa_1 = k\sqrt{m_1 + M}$, $M$ is the mass of the Sun and $n_1$ stands for the mean motion [2]. Each canonical pair obeys the canonical equations

$$\frac{d\boldsymbol{\Gamma}_1}{dt} = \frac{\partial R_{12}}{\partial \gamma_1}, \qquad \frac{d\gamma_1}{dt} = -\frac{\partial R_{12}}{\partial \boldsymbol{\Gamma}_1} \tag{5}$$

[2] which relates to the semimajor axes $a_1$ via the $3^{rd}$ Kepler law $n^2 a^3 = \kappa^2$

where $\boldsymbol{\Gamma}_1 = (L_1, H_1, G_1)^{\mathrm{T}}$ has the conjugate vector $\boldsymbol{\gamma}_1 = (l_1, h_1, g_1)^{\mathrm{T}}$. As an example we now show how the perturbations act on the Delaunay element $H_1$. To derive the respective perturbations of the first order we need to compute $\delta H_1 = \int \frac{\partial R_{12}}{\partial h_1} dt$. Inserting the perturbing function (3) one can construct the partial derivative with respect to the conjugate variable

$$\delta H_1 = m_2 \int \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} \frac{\partial C_{\mathrm{j_1,j_2}}}{\partial h_1} \cos[(j_1 n_1 + j_2 n_2)t + D_{\mathrm{j_1,j_2}}]dt, \qquad (6)$$

which integrates to

$$\delta H_1 = m_2 \sum_{j_1=-\infty}^{\infty} \sum_{j_2=-\infty}^{\infty} \frac{\partial C_{\mathrm{j_1,j_2}}}{\partial h_1} \frac{\sin[(j_1 n_1 + j_2 n_2)t + D_{\mathrm{j_1,j_2}}]}{(j_1 n_1 + j_2 n_2)} \qquad (7)$$

and hence so-called *small divisors* appear. Note that the $C_{\mathrm{j_1},j_2}$ depend only on the action variables $\boldsymbol{\Gamma}$ and the $D_{\mathrm{j_1},j_2}$ only on the angle variables $\boldsymbol{\gamma}$. Whenever the two mean motions involved, $n_1$ and $n_2$, are in resonance, the small divisor – being a number close to zero – makes the perturbation for that element very large. A good example is the *great inequality* between Jupiter and Saturn. The two giant planets are in the *mean motion resonance* 5:2 which means that $2n_{\mathrm{jup}} - 5n_{\mathrm{sat}} \sim 0$ for the summation indices $j_1 = 2$ and $j_2 = 5$. Inserting the values of the mean motions for Jupiter and Saturn one gets in degrees per day $2 \cdot 0.^{\circ}08309 - 5 \cdot 0.^{\circ}03346 = -0.^{\circ}00112$ and thus for the great inequality a period of $\sim 880$ years. A more detailed computation shows that the amplitude of this perturbation in longitude reaches values up to almost $0.^{\circ}5$ for Jupiter due to Saturn (and vice versa for Saturn due to the perturbation of Jupiter almost $1^{\circ}$)[3].

### 2.2 High Order Resonances, Secular Resonances and the Fundamental Frequencies

In the summation of the Fourier expansion of the perturbing function only a finite number of terms for $j_1$ and $j_2$ are to be taken into account, because beyond these terms the amplitudes are rather small even for "small divisors". The main point is, that it is known since Poincaré [23], that the series expansion of the perturbation function is not convergent for high orders and has to be truncated at some point. Nevertheless high order resonances (like $j_1 = 29$ and $j_2 = -72$ for Jupiter and Saturn) or secular resonances (see below) may appear inside the main resonance and occasionally produce chaotic motion, due to overlapping separatrix layers for larger eccentricities. This is less important for planetary motion than for the motion of the asteroids. Of special interest for the stability of the planetary system are long-period perturbations, which can be computed by averaging over the fast variable $l = M = nt$. The system described in (5) then decouples into two sets of equations, where one involves only the eccentricity and

[3] The necessary double integration for the element $l_1$ leads to a divisor of the form $(j_1 n_1 + j_2 n_2)^2$ and makes the perturbation especially large.

the perihelion, while the other involves only inclinations and nodes (there are no equations for $L$ and $l$). The solutions of the so–called Lagrange system describe the motion of the nodes and the perihelia of the orbits. The determination of the secular frequencies is of great interest in connection with the chaotic behaviour of the planetary system (see Sect. 4.2).

When one takes into account that the nodes and the perihelia of the orbits change slowly with time ($\omega = \omega_o + \omega_1 t$ and $\Omega = \Omega_o + \Omega_1 t$) the aforementioned "phase coefficient" in (3) becomes also time dependant. Therefore, in the integration of the secular evolution another small divisor, (e.g. $j_1\omega_1 + j_2\omega_2 \approx 0$) may appear which leads again to strong perturbations due to a *secular resonance.*

In the so–called *Kozai resonance* [12] the eccentricities and the inclinations of small bodies perturbed by the planets are coupled in the sense that the eccentricity of the orbit has a maximum when the inclination has a minimum and vice-versa. This can be explained by the fact that for a constant semimajor axis $a$ the Delaunay element $H = \kappa\sqrt{a(1-e^2)}\cos i$ is a constant of motion (e.g. [20]).

The recently discovered *three-body mean motion resonances* take into account in the series expansion (3) the mean motions of the perturbed body and two perturbing planets ($j_1 n_1 + j_2 n_2 + j_3 n_3 \approx 0$), and seem to act especially in the motion of asteroids [20]. Without going into the details we point out that overlap of these resonances can lead to chaotic motion (for an extensive description see [4]).

## 3 Dynamics of Comets and Asteroids

Our solar system is populated by a large number of small bodies orbiting the Sun in more or less eccentric orbits. The distinction between the two physically different populations – the comets and the asteroids – is already evident because of the different apparition for the observer on Earth. We can divide the comets, according to their orbits, into short–period comets with orbital periods smaller than 200 years and the long–period ones. The latter very probably come from the Oort Cloud, where on the order of $10^8$ comets may be present with aphelion distances up to 0.4 pc. Their hyperbolic orbits, which are not confined to small inclinations, led them to the inner part of the solar system due to perturbations from passing stars, interstellar clouds and galactic tides. Within the group of short–period comets[4] we can distinguish the Jupiter family with periods less than that of Jupiter and the ones with orbits similar to that of comet Halley [5]. The orbital inclinations of comets of the Jupiter family are small and most of them are supposed to come from the second reservoir for comets, the *Edgeworth-Kuiper* belt outside Neptune's orbit. The orbits of short-period comets are often deformed due to close encounters with planets, particularly with Jupiter.

[4] we now know about 150 of this group

[5] with the orbital elements $a = 17.94\ AU, e = 0.967, i = 162.^o2, \omega = 112^o, \Omega = 58.^o1$ and $T = 1986\ 02\ 19.0$

These close encounters can be modeled via scattering and are the source of unpredictability of the orbit for longer time intervals. Using numerical integrations of real and fictitious objects [5] and simplified mapping methods [3] the statistical properties for these encounters were determined and the respective Poincaré surfaces of sections unveiled the chaotic structure of these orbits. In Fig. 1 we see the orbits of comet Halley for two slightly different initial conditions, which lead, after several close approaches to Jupiter, to completely different dynamical behavior. Whereas the upper orbit leads to hyperbolic escape after $2.10^5$ years, the lower orbit shows the typical character of jumping from one mean motion resonance with Jupiter to another. In the lowest graph we can see that the comet's orbit is in the Kozai resonance, where the inclination and the eccentricity move oppositely (well visible especially between 170 and 190 kyrs).

We can also estimate dynamical lifetimes of these objects [19], which are on the order of $10^6$ years. This lifetime is determined by: (i) collision with a planet (e.g. crash of SL9 on Jupiter in 1994), (ii) breakup in a sun-grazing encounter or (iii) escape from the solar system after a close encounter with a giant planet.

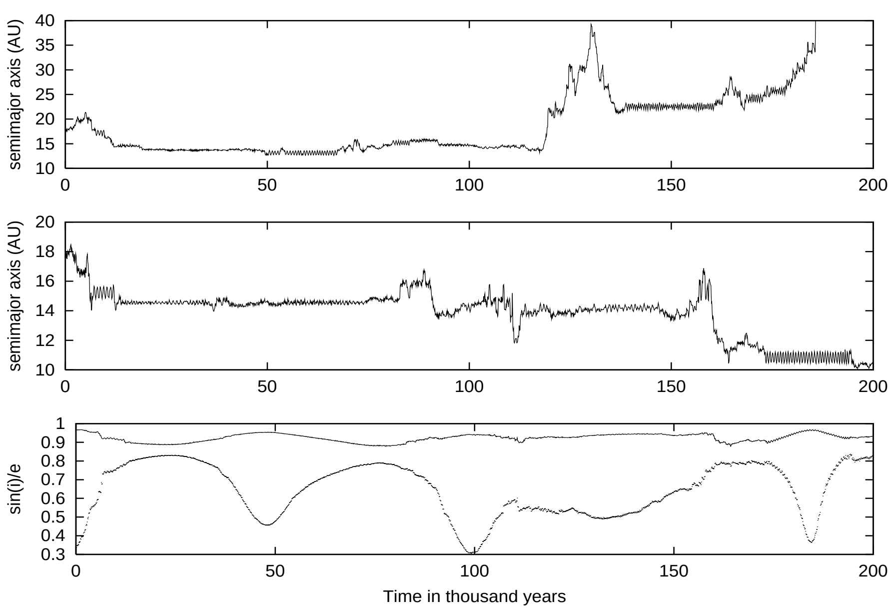



**Fig. 1.** The time evolution for two fictitious objects in Halley like orbits with a $\Delta M = 0.^\circ 01$. The upper graphs plot the semimajor axis versus the time, the lower graph shows the eccentricity (upper curve) and the sin($i$) (lower curve) to illustrate that the $2^{\mathrm{n}d}$ object is in the Kozai resonance

For the asteroids we can distinguish 4 different groups[6] (i) the Edgeworth-Kuiper-Objects (KBO, moving outside Neptune) (ii) the cloud of Jupiter Trojans (moving close to the Lagrange equilibrium points $L_4$ and $L_5$), (iii) the main belt asteroids (with semimajor axes between Mars and Jupiter) and (iv) the Near Earth Asteroids (NEAs), with orbits which bring them close to the Earth. Only the Trojans form a well defined group, the other ones are not so well separated, respectively they change their membership to a specific group.

The members of the NEAs are usually divided into three subgroups: the ATENS, with a semimajor axis smaller than that of the Earth and an aphelion distance $Q = a(1 + e) > 0.983$ AU (mean perihelion distance of Earth), the APOLLOS, with a semimajor axis larger than that of the Earth and a perihelion distance $q = a(1 - e) < 1.017$ AU (mean aphelion distance of Earth) and the AMORS, with a semimajor axis larger than that of the Earth and a perihelion distance $1.017 < q < 1.3$ AU (they do not cross Earth's orbit, but they stay inside the orbit of Mars)

The number of known asteroids is growing rapidly because during the observing programs of the NEAs more and more small and also large objects[7] are discovered. The actual numbers of asteroids in addition to the $\approx 20000$ main belt asteroids – are Atens (172), Apollos (1043), Amors (1013), Centaurs (125), Jupiter Trojans around the preceding Lagrange point $L_4$ (962), $L_5$ (602) and KBOs (664).

In Fig. 2 one can see the sculpting of the main belt inside Jupiter's orbit which is mainly due to mean-motion resonances, as well as three-body resonances and secular resonances (inner main belt) and also by the chaotic behavior of separatrix crossing (e.g. 3:1 mean motion resonance). We can order the importance of the acting resonances [22] as follows: (1) three–body resonances with Jupiter and Saturn, (2) resonances with Mars, (3) mean motion resonances with Jupiter, and (4) three–body resonances involving both Mars and Jupiter. In the inner part of the main belt Mars is – besides the secular resonances with Jupiter and Saturn [21] – an important perturbing planet; the relatively large eccentricity of its orbit compensates for its small mass. The middle part $2.5 \leq a \leq 3\ AU$ is the most stable one. In the outer belt Jupiter is acting with its resonances and sub–resonances.

An important step further in the knowledge of solar system dynamics occurred, when Wisdom [25] discovered that the depletion of the main belt in the 3:1 mean motion resonance with Jupiter is caused by the chaotic behaviour of asteroids located there. The asteroid suffers from a sudden increase in the eccentricity which may bring it into a resonance overlap area, which in turn leads to an additional increase in eccentricity and finally to a close approach with Mars. This process of separatrix crossing is a basically chaotic behavior known

[6] we exclude the Centaurs as a group which are asteroids outside Jupiter but inside the 3:2 resonance with Neptune

[7] Quaoar, a recently discovered KBO, has a semimajor axis a=43.4 AU, an eccentricity of e=0.03 and a diameter of 1250km (which is half the size of Pluto).

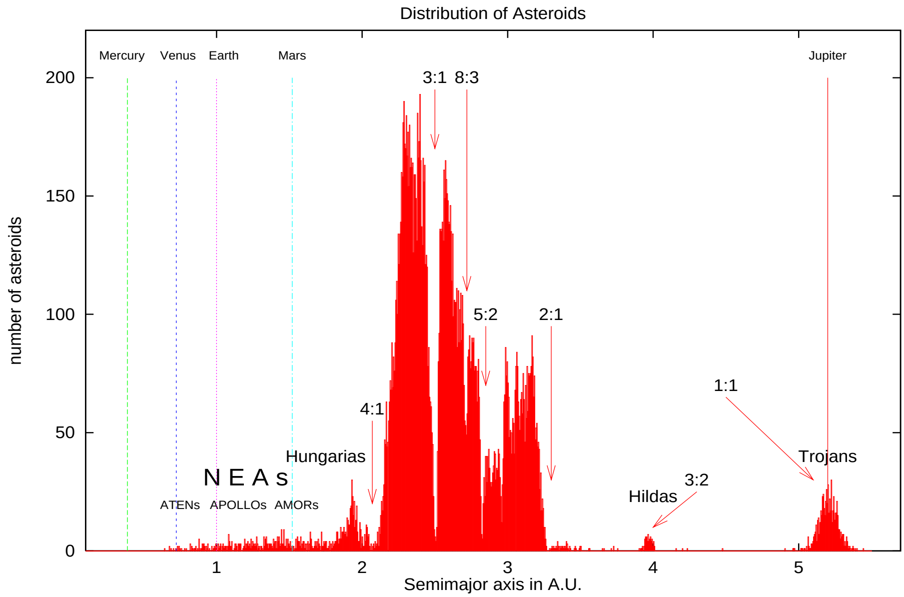



**Fig. 2.** Distribution of asteroids inside Jupiter's orbit

from the perturbed pendulum, where the qualitative change from libration to circulation is due to the sensitivity on the initial conditions.

The riddle of the different dynamical behaviour of the 2:1 resonance, which is almost depleted from planetoids, and the 3:2 Hilda family of asteroids has been solved recently by the work of different authors (e.g. Ferraz-Mello and Nesvorny [6]) and is primarily due to an early depletion of asteroids in the 2:1 resonance because of the different location of Jupiter and Saturn in the early days of the solar system, where they were somewhat closer. A detailed description of the structure of these two resonances can be found in the previously mentioned book by Morbidelli [20] (p. 286–294).

In Fig. 3 one can see how a NEA jumps from one mean motion resonance with the Earth to another (upper panel), which is caused by more or less close encounters with this planet (lower panel). The detailed transport of asteroids from the Kuiper belt to the Centaurs and from there to the inner regions of the solar system as well as the transport from the main belt to Earth–crossing orbits, is governed mainly by secular resonances [24].

## 4 The Planets

The problem of the stability of our planetary system was first studied by Laplace in the $18^{th}$ century, who found that the semimajor axes of the planets suffer only periodic changes up to the first order; 50 years later Poisson extended this theorem up to second order. As already mentioned Poincaré [23] discovered

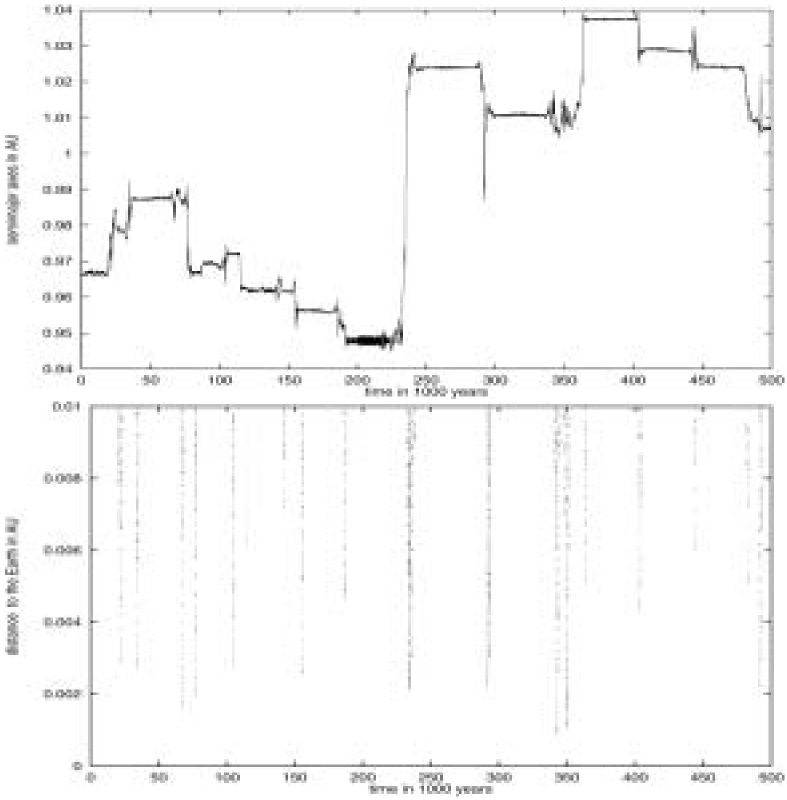



**Fig. 3.** The jumping orbit of 2262 Aten

that the usual perturbation techniques used to represent the solution in form of power series in small parameters such as the eccentricities, inclinations and/or the planetary masses are not convergent, because of small divisors. Using the method of averaging over the mean motions of the planets (and thus not being able to determine the longitude of the planet) one can solve the secular part of the perturbing function. To lower orders one can find good approximations to the solutions because the system decouples into an inclination–node independent and an eccentricity–perihelion independent system. For a qualitatively good solution over millions of years one can introduce high order terms which result in a shifting of the proper mode frequencies and combinations of them [1]. Using numerical techniques we can nowadays – with the aid of very fast computers – integrate the full equations of motion up for times scales comparable to the age of the solar system [10].

### 4.1 Numerical Solutions

We have integrated the full equations of motion for a time interval of 200 million years ($\pm 100$ million years) for a dynamical model consisting of the Sun and 8

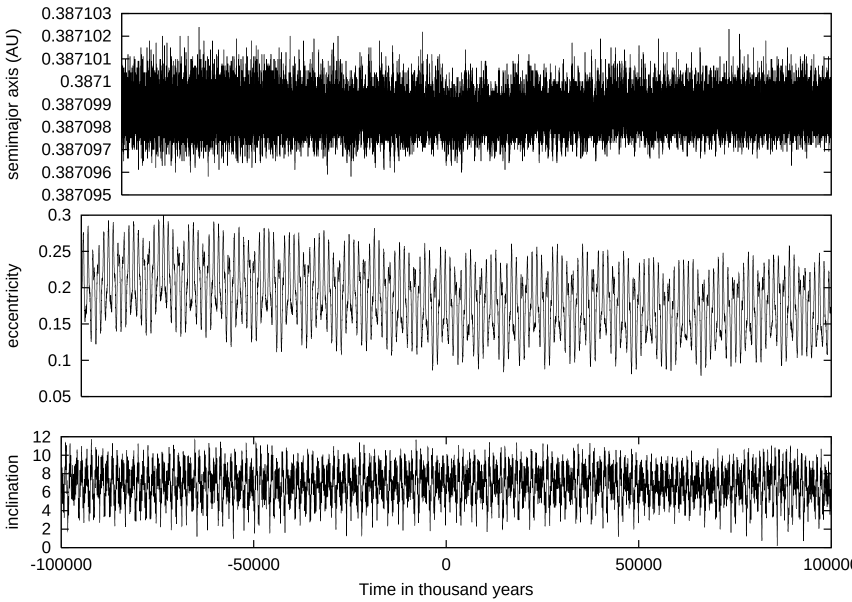



**Fig. 4.** Semimajor axis, eccentricity and inclination of Mercury for 200 million years

planets (treated as point masses, Earth + Moon as one body, relativistic effects added afterwards) with the Lie series method ([18], [8]). In Fig. 4 one can see the relative large variations in the eccentricity and inclination for Mercury, while the semimajor axis is very constant even for such a long integration. These results illustrate the precision of the method because the first sign of the lack of precision of an integration is a secular drift of the semimajor axes of the innermost planet in a simulation of the motion of the planetary system [7].

The coupling between Earth and Venus, which have almost the same mass, is visible in Fig. 5. We show – besides the time interval -2.5 to +2.5 million years around today – the time evolution of their eccentricities for 5 million years in the past and in the future. It is remarkable that no differences at all can be seen there!

For the interval of 200 million years the minimum and maximum values of the semi-major axes, the eccentricities and the inclinations of the planets are shown in Table 1.

### 4.2 Determination of the Fundamental Frequencies

The analysis of the data to determine the so-called fundamental frequencies (the motion of the nodes and of the longitudes of the perihelia) was done using a program provided by [2] which uses a Chebycheff approximation of the data, determines the largest amplitude, subtracts it and repeats the analysis (etc.)

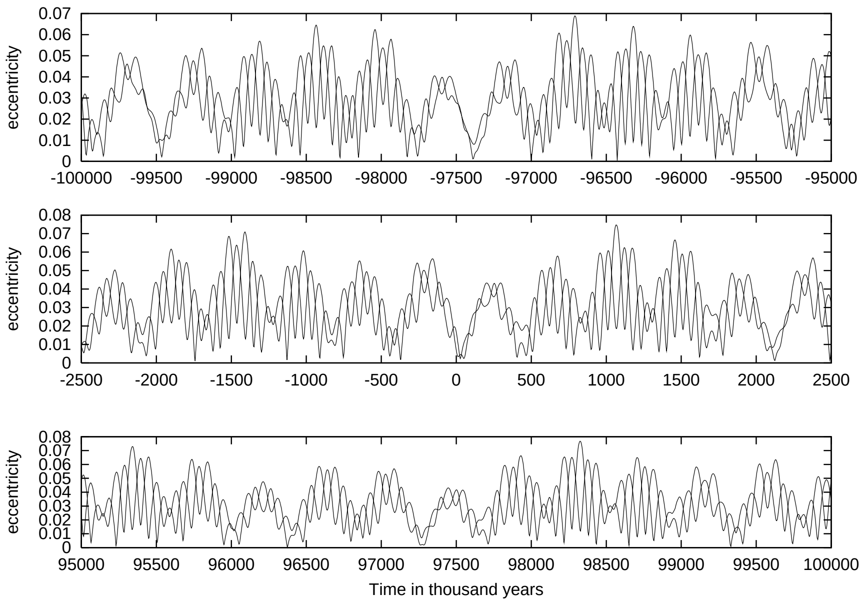



**Fig. 5.** Eccentricities of the Earth and Venus

**Table 1.** Extreme values of the action variables for $2.10^8$ years

| planet | $a_{min}$ | $a_{max}$ | $e_{min}$ | $e_{max}$ | $i_{min}$ | $i_{max}$ |
|---|---|---|---|---|---|---|
| Mercury | 0.38710 | 0.38710 | 0.07874 | 0.29988 | 0.17600 | 11.72747 |
| Venus | 0.72332 | 0.72336 | 0.00002 | 0.07709 | 0.00076 | 4.91515 |
| Earth | 0.99997 | 1.00004 | 0.00002 | 0.06753 | 0.00075 | 4.49496 |
| Mars | 1.52354 | 1.52386 | 0.00008 | 0.13110 | 0.00291 | 8.60320 |
| Jupiter | 5.20122 | 5.20504 | 0.02513 | 0.06191 | 1.09172 | 2.06598 |
| Saturn | 9.51281 | 9.59281 | 0.00742 | 0.08959 | 0.55867 | 2.60187 |
| Uranus | 19.09807 | 19.33511 | 0.00008 | 0.07835 | 0.42170 | 2.73888 |
| Neptune | 29.91013 | 30.32452 | 0.00001 | 0.02317 | 0.77977 | 2.38597 |

This new determination ([7]) is compared to the values published by previous authors in Table 2. One can see that these frequencies are quite close to the already published ones using other methods.

Laskar [13] used a different approach to model the long term evolution of the planetary orbits: he integrated numerically the secular system, (truncated up to $2^{\text{nd}}$ order in the masses and to $5^{\text{th}}$ order with respect to the eccentricities and the inclinations). The surprising result was the discovery that a secular resonant term, namely $\theta = 2(\dot{\varpi}_{\mathrm{M}} - \dot{\varpi}_{\mathrm{E}}) - (\dot{\Omega}_{\mathrm{M}} - \dot{\Omega}_{\mathrm{E}})$, where E stands for the Earth and M stands for Mars, is alternatively librating and circulating [15]. This clear

**Table 2.** Fundamental frequencies for the motion of the perihelion in $''$ per year after the Lagrange-Laplace theory (LLT), the theory of Bretagnon (B84) [1], the theory of Laskar (NGT) [14] and a new determination (NEW)

| | LLT | B84 | NGT | NEW |
|---|---|---|---|---|
| $g_1$ | 5.4633 | 5.6136 | 5.5689 | 5.6276 |
| $g_2$ | 7.3477 | 7.4559 | 7.4555 | 7.4441 |
| $g_3$ | 17.3283 | 17.2852 | 17.3769 | 17.5668 |
| $g_4$ | 18.0023 | 17.9025 | 17.9217 | 17.9373 |
| $g_5$ | 4.2959 | 4.3080 | 4.2489 | 4.2567 |
| $g_6$ | 27.7741 | 28.1483 | 27.9606 | 28.2445 |
| $g_7$ | 2.7193 | 3.1534 | 3.0695 | 3.0468 |
| $g_8$ | 0.6333 | 0.6735 | 0.6669 | 0.6727 |

indication of chaos does not a priori mean that the planetary system will be unstable. Nevertheless these important results show that the orbital elements of the planetary system – especially the terrestrial planets – lie in a thin chaotic layer in phase space.

In a further step Laskar [14] numerically integrated the averaged equations of motion over even longer time scales (several $10^9$ years with a time step of 250 years for this integration). The most interesting results are shown in Fig. 6 where the development of the eccentricities of the innermost three planets, respectively the maximum and minimum value over a time span of some billion years is plotted. The two lines on the bottom with quite a similar behaviour belong to Venus (respectively to the Earth) and show the dynamical coupling of these two planets. For Mercury we see a line of minimum values and a line of maximum

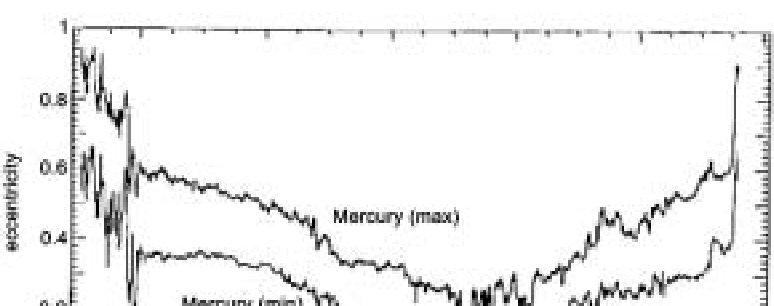



**Fig. 6.** Evolution of the eccentricities of Venus, Earth and Mercury from -6.6 Gyrs to 3.5 Gyrs (after [15])

values which are parallel (separated by $\Delta e \sim 0.15$) over the whole time span of $10^{10}$ years. The most interesting fact is that Mercury's orbit seems to be able to achieve values of $e$ close to unity! Laskar concludes that this could lead (in the future, because the system is symmetric with respect to the time) to a close encounter with Venus. The most important point of criticism is that the development of the perturbation function is not convergent for large inclinations and eccentricities! Thus the solution which he computes does not describe the dynamical development of the planetary system. Nevertheless Laskar argues that the secular system, which keeps the semimajor axes of the planets constant, is even less chaotic than the real one and additional degrees of freedom" will probably lead to even stronger chaotic behaviour, as in general, addition of degrees of freedom increase the stochasticity of the motion" ([15]).

Because of the shortcomings in the equations of motions these computations cannot give a conclusive answer to the question of the stability of the planetary system, but there is no doubt that the motion of the planets is NOT regular (and not quasiperiodic). This fact is mainly due to Mercury and Mars in the inner solar system; Earth and Venus are strongly coupled in their dynamical behaviour and the outer planets show regular behaviour in their motions (with the exception of Pluto, which is a Kuiper belt object moving in a 2:3 resonance with Neptune).

The results discussed in this chapter raise several questions:

• is the sudden increase of Mercury's orbital eccentricity (in short time scales of $10^5 - 10^6$ years, compare Fig. 6) due to the shortcomings of the equations of motions, or is it a real phenomenon?

• what is Mercury's action on the other planets when its orbit is highly eccentric?

• what is the possible outcome of a close encounter between Mercury and Venus?

To clarify this question we have undertaken numerical integrations of the full equations of motion in different models: only the inner planets, the inner planets plus Jupiter, the inner planets plus Jupiter and Saturn and the complete planetary system.

### 4.3 Mercury as Perturber of the Inner System

Figure 7 shows the evolution of the eccentricity of Mercury for 1 million years, starting with a large eccentricity $e_{\text{ini}} = 0.95$ in 2 different models, namely with and without Jupiter: in the first model after only $10^5$ years the eccentricity can drop from 0.95 to 0.6 and then rise again (upper panel)! Thus the result of Laskar is confirmed that a "sudden" increase of Mercury's eccentricity is a possible scenario. On the lower panel of Fig. 7 we depict the behaviour in a model with Jupiter, and here one can see that, although a close encounter occurred, the eccentricity of Mercury did not drop below $e = 0.75$. Other test calculations confirm the result, that the outer planetary system stabilizes the inner one (a longer paper with computations for different dynamical models is in preparation). In all our computations the consequences of close encounters

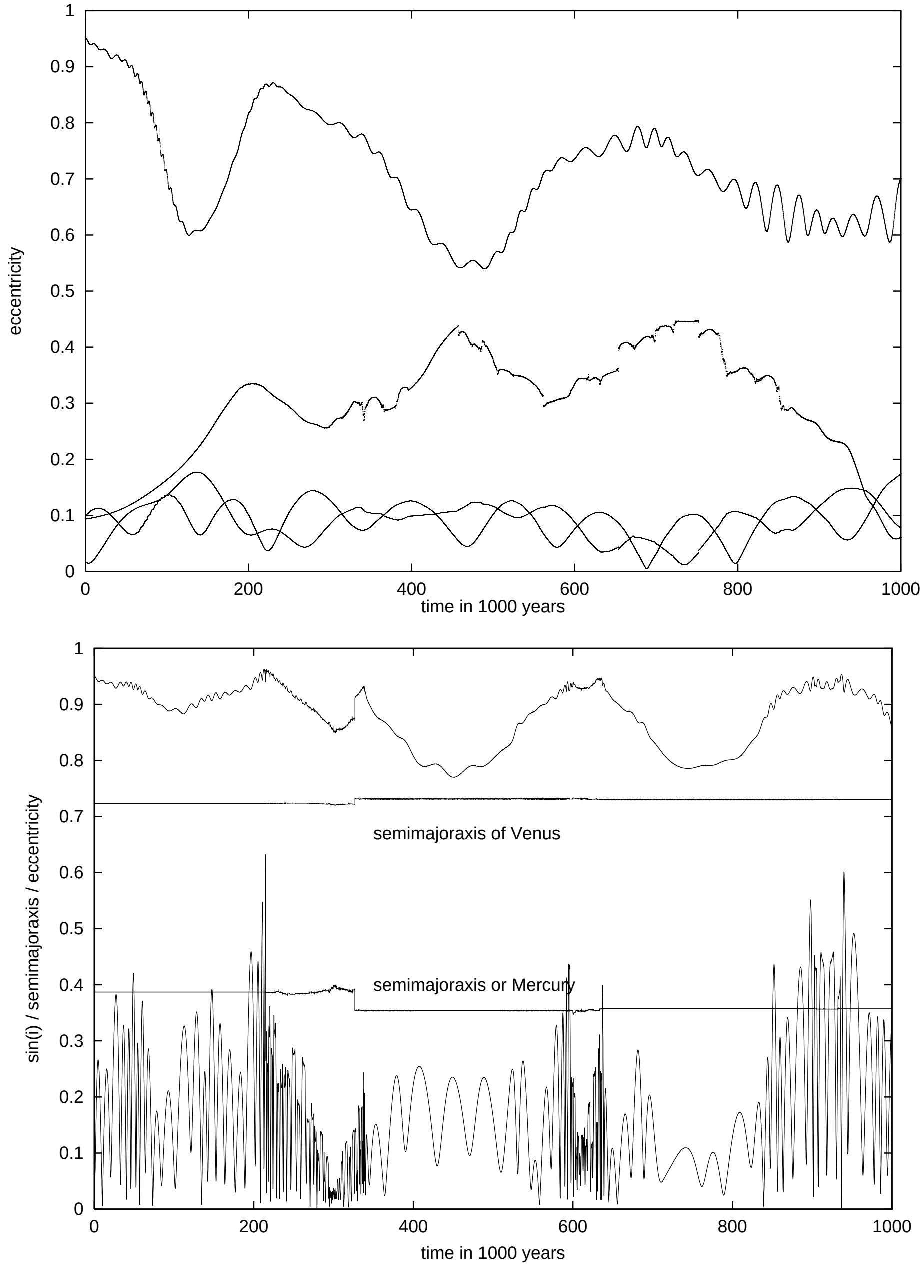



**Fig. 7.** Eccentricities of the inner planets for 1 million years in a model without the outer planets; the two bottom lines are for Venus and Earth, the upper one for Mercury and the one in the middle for Mars (upper panel). Eccentricity, semimajor axis, and $\sin i$ of Mercury and semimajor axis of Venus in a model with Jupiter for 1 million years (lower panel)

between Venus and Mercury never resulted in an escape, but only in a small change in the semimajor axis. We therefore conclude that escape can only happen after several cascade–like encounters, and then it may be that Jupiter could throw Mercury far out into the Kuiper belt! Regarding the action of the unstable orbit of Mercury on the other planets, it turned out that the influence is small, owing to the small mass of this planet. Only an unlikely close encounter of Mercury (or quasi–collision, which in fact we observed in one of our experiments) could lead to larger changes of the orbital elements.

### 4.4 The Obliquity of Planets

The results of an investigation of the obliquities [16] of the planets were quite surprising: with one exception all inner planets are – in what concerns the obliquity of their rotation axes – in a mode of chaos. That means that within relatively short time scales of millions of years their axes can vary substantially in space. For Mars the amplitude of this chaotic development of the obliquity is 15° within some hundred thousand years and in 50 million years it may gradually evolve from almost 60° to 15°. At the same time the precession rates also vary by a factor of two. Although Venus now possesses an orientation of the rotational axis very close to that of our Earth, the obliquity could reach a value close to 90° and therefore it would almost "roll" in its orbital plane (like Uranus is actually doing it). For the Earth the situation is quite different: in an 18 Myrs integrations of the precession equations it was found that the precession rate can change to a negative precession of $p = -39''/\mathrm{y}r$. The actual values for the Earth of $\epsilon_o = 23°$ and $p = 50''/\mathrm{y}r$ lie well inside a quiet region with only very small changes for the maximum and minimum values. For $55° < \epsilon < 90°$ there exist a large chaotic region allowing variations in the order of 35°. The same integrations were undertaken without the presence of the Moon and the obliquity turned out to lie in a region of large chaos (such as Mars is actually in). Thus it seems that only the presence of the Moon stabilizes the obliquity so that only small changes in $\epsilon$ occur and occurred in the last millions of years!

## 5 Conclusions

Let us summarize the most interesting manifestation of chaos in our planetary system:

• all small bodies in our solar system – asteroids and comets – suffer from chaotic motion either directly due to close encounters or after long term evolutions

• there is a thin layer of chaos in phase space, where the inner planets move

• on a long time scale diffusion may bring the planets into a state of larger chaos which may allow for large eccentricities of Mercury. The scenario which follows then is still not known and seems to be more speculation than outcome of scientific investigations; at any rate an immediate escape of a planet does not seem possible. A snowball like effect is more probable, where Mercury changes

Venus' orbit, Venus comes close to the Earth and changes its orbit, the Earth itself is then perturbing Mars, which now may achieve larger eccentricities ...

• Jupiter and Saturn keep the inner planets in stable orbits for quite long times

• large changes in the orientation of the obliquity of the planets are possible due to chaotic motion of the inner planets. On the contrary the Earth's spin axis is stabilized by the Moon which turns out to be in a nonchaotic region of configuration space.

## 6 Epilog

The knowledge of more than 100 planets orbiting other stars has opened new interest for the dynamics of our own planetary system, where it is difficult to detect the weak chaoticity the planetary orbits are in. This is different for extrasolar planets: in a recent work it was found that the planets in HD12661 are apparently in a chaotic region of phase space [11]. In Fig. 8a we show the time evolution of the semimajor axis of the inner planet, where it jumps from one high order mean motion resonance to another close by. In numerical experiments of the dynamical evolution of our planetary system with fictitious larger masses of the terrestrial planets a similar behaviour was found (Fig. 8b): taking for Mars the mass of the Earth leads to a chaotic behaviour of the semimajor axis of the

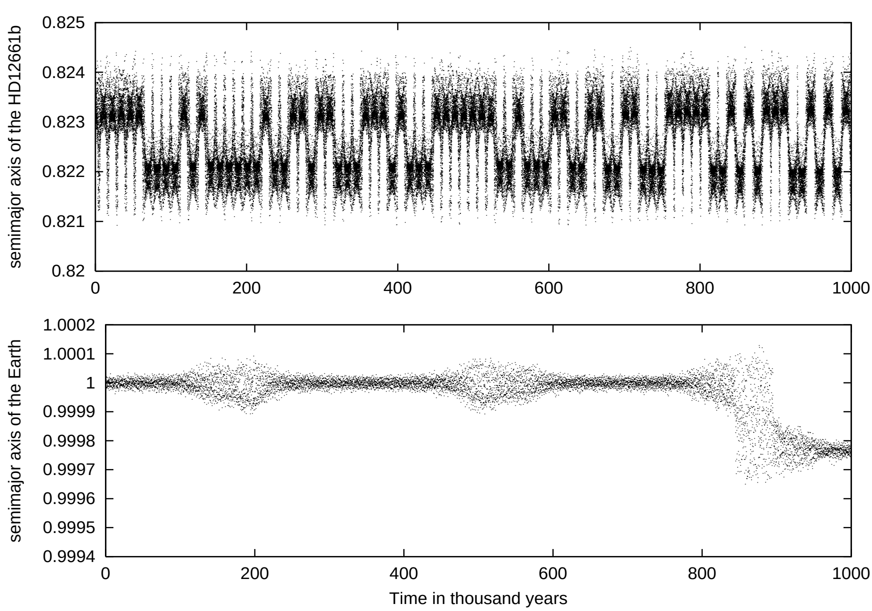



**Fig. 8.** Time evolution for 1 million years of the semimajor axis of the inner planet of HD12661 (upper panel) and of the Earth (in a model with a fictitious mars with 10 time its actual mass) (lower panel)

Earth (the masses of the other planets were unchanged). Because we can expect many more discoveries of planets even with masses comparable to our terrestrial planets in the future, the knowledge of our own system is a basis for the future research in the dynamics of other planetary systems.

**Acknowledgements**

For improving the paper I have to thank primarily J. Howard (Boulder). F. Freistetter and the graduate students of the Astro Dynamics Group in Vienna helped to finalize the manuscript. This work was partly carried out within the framework of the FWF project P14375-TPH.

## References

1. P. Bretagnon, P.: Cel. Mec. Dyn. Astron. **34**, 193 (1982)
2. J. Chapront: Astron. Astrophys. **109**, 181 (1995)
3. R.V. Chirikov, V.V. Vecheslavov: Astron. Astrophys. **221**, 146 (1989)
4. G. Contopoulos: *Order and Chaos in Dynamical Astronomy* (Springer–Verlag, Berlin Heidelberg New York 2002)
5. R. Dvorak, J. Kribbel: Astron. Astrophys. **227**, 264 (1990)
6. S. Ferraz-Mello, D. Nesvorny: Astron. Astrophys. **320**, 672 (1997)
7. C. Gamsjäger: *Eine Neubestimmung der Basisfrequenzen in den Planetenbewegungen*, Master Thesis, University of Graz (2002)
8. A. Hanslmeier, R. Dvorak: Astron. Astrophys. **132**, 203 (1984)
9. M. Hénon, C. Heiles: AJ **69**, 73 (1964)
10. T. Ito, K. Tanikawa: MNRAS **336**, 483 (2002)
11. L. Kiseleva-Eggleton, E.Bois, N. Rambaux, R. Dvorak: ApJ **578L** 145 (2002)
12. Y. Kozai: AJ **67** 591 (1962)
13. J. Laskar: Icarus **88** 266 (1990)
14. J. Laskar: Astron. Astrophys. **287** L9 (1994)
15. J. Laskar: Cel. Mec. Dyn. Astron. **64**, 115 (1996)
16. J. Laskar, P. Robutel: Nature **361**, 608 (1993)
17. M. Lecar, F.A. Franklin, M.J. Holman, N.J. Murray: Ann.Rev. of Astron. and Astrophys. **39**, 581 (2001)
18. H. Lichtenegger: Cel. Mech. **34**, 357 (1984)
19. E. Lohinger, R. Dvorak, C. Froeschlé: Earth, Moon and Planets **71** 225 (1995)
20. A. Morbidelli: *Modern Celestial Mechanics : Aspects of Solar System Dynamics* (Taylor & Francis, London 2002)
21. P. Müller, R. Dvorak: Astron.Astrophys. **300** 289 (1995)
22. D. Nesvorný, A. Mordbidelli: AJ **116** 3029 (1998)
23. H. Poincaré: *Les Méthodes Nouvelles de la Mécanique Céleste*, tome I–III, (Gauthier–Villard et fils, Paris 1892)
24. P. Robutel, J. Laskar: Icarus **152** 4 (2001)
25. J. Wisdom: Icarus **56**, 51 (1983)

# Dynamics of Extrasolar Planetary Systems: 2/1 Resonant Motion

John D. Hadjidemetriou and Dionyssia Psychoyos

University of Thessaloniki, Department of Physics, 541 24 Thessaloniki, Greece

**Abstract.** A systematic study of the dynamics of resonant planetary systems is made, based on the existence and stability character of families of periodic orbits of the planetary type. In the present study we consider planetary systems with two planets, moving in the same plane. We explore the whole phase space close to the 2/1 resonance, for the masses of the observed planetary system HD82943. We find four basic resonant families of periodic orbits at the 2/1 resonance, and show that large regions on the families correspond to stable motion, even for large values of the eccentricities of the two planets and for intersecting planetary orbits. The initial phase of the two planets plays a crucial role on the stability of the system. It is close to a periodic orbit that stable motion of a planetary system can exist. So, the study of the families of periodic orbits provides a systematic way to find all the regions of phase space where a resonant planetary system could exist in nature. Planetary systems with large eccentricities can exist in nature only if they are close to a resonance. Indeed, we show that the real planetary system HD82943 is close to a stable periodic orbit. The alignment of the line of apsides of the planetary orbits plays also a stabilizing role.

## 1 Introduction

The study of extrasolar planetary systems is an important new field of research in dynamical astronomy, following the discovery during the last decade of planetary systems around distant stars. A complete catalogue of extrasolar planetary systems can be found in the web site *http://www.obspm.fr/encycl/catalog.html*, maintained by Jean Schneider. There are 91 confirmed extrasolar planetary systems, with ten of them having two planets and two having three planets. In some cases, the eccentricities of the two planets are quite large, and their masses are comparable to the mass of Jupiter. In all these cases, the two planetary orbits are in a mean motion resonance.

There are several problems associated with the extrasolar planetary systems. First are the cosmogonic problems: how were these systems formed? Another problem is the stability of such systems. What makes them stay there, for millions or billions of years? These systems are very different from our own solar system, because the planetary eccentricities are large, but they do exist in nature. What mechanism keeps them stable? A third question refers to the possibility of existence of life is an extrasolar planetary system. This implies the existence of Earth-like planets, i.e. planets of the size of the Earth, in nearly circular orbits at a distance of about 1AU from the Sun. The observational techniques at the moment are not so accurate and consequently such planets cannot be observed.

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 412–432, 2003.

Only larger planets are detected at present. But knowing the position of the larger planets, we can study the stability of a fictitious small planet at about 1AU, and thus find if such a planet could possibly exist. A study on this problem is in Celletti et al. (2002), Jones and Sleep (2002).

In this paper we address the problem of the stability of an extrasolar planetary system. There are several studies on the stability of the observed extrasolar planetary systems, based on numerical integrations, starting with initial conditions close to those corresponding to the observed values: (Beauge et al. 2003, Callegari et al. 2002, Ferraz-Mello 2002, Ford et al. 2001, Gozdziewski et. al. 2002, Kinoshita and Nakai 2001a,b, Kiseleva et.al. 2002, Laughlin and Chambers 2001, Lissauer and Rivera 2001, Malhotra 2002a,b, Murray et al, 2001, Peale and Lee 2002, Rivera and Lissauer 2001). In the present study we propose a new approach to detect stable motion. A systematic method is presented, based on periodic orbits, to find all the regions of the phase space where stable motion exists and consequently a real planetary system could be found in nature. A study on the dynamics, based on families of periodic orbits, is made in Hadjidemetriou (2002), for the 2/1 and 3/2 resonant systems.

We consider an extrasolar planetary system with two planets, moving in the same plane, taking into account the gravitational attraction between the two planets. This is a special case of the general planar three body problem. The study is based on the existence of periodic orbits in a rotating, synodic, frame, because it is the periodic orbits and their stability character, that determine critically the structure of the phase space. So, the knowledge of all the basic families of periodic orbits in a dynamical system will provide useful information on the evolution of the system. In particular, the position of the stable periodic orbits in phase space gives a systematic way to find those regions where stable motion can exist. This is so, because it is only close to a stable periodic motion where a real planetary system can exist. In this way, we are able to find all the resonant planetary systems that could exist in nature.

As we shall explain in the next section, there are two types of periodic orbits of the planetary type (in the rotating frame): Periodic orbits where the two planets move in circular orbits and periodic orbits where the two planets move in elliptic orbits (slightly distorted due to the gravitational interaction between the two planets). These latter periodic orbits are necessarily resonant, i.e. the ratio of the periods of revolution of the two planets is a rational number. The eccentricities of the two planets vary along such a resonant family of periodic orbits, starting from zero values and reaching high values, while the semimajor axes remain almost constant.

An extrasolar planetary system, even with large masses and eccentricities, can be stable, provided that it is close to a stable resonant periodic orbit. Indeed, all the extrasolar planetary systems with large eccentricities that have been observed, are close to a resonance. In particular, we study the 2/1 resonant case, and we find that stable families of resonant periodic orbits exist, with large eccentricities of the planets. As a model system we consider the observed extrasolar planetary system HD82943. In this system two planets have been observed (Israelinian et al. 2001): The masses of the two planets, expressed in Jupiter masses, are $m_1 \sin i = 0.88J$, $m_2 \sin i = 1.63J$, the semimajor axes are $a_1 = 0.73\text{AU}$, $a_2 = 1.16\text{AU}$, their periods are $T_1 = 221.6\text{d}$, $T_2 = 444.6\text{d}$, their eccentricities are $e_1 = 0.54$, $e_2 = 0.41$ and the mass of the sun is $m_{sun} = 1.05$ solar masses. We clearly see that this is a resonant planetary system, with 2/1 mean motion resonance of the planets, $T_2/T_1 = 2.006$.

We computed four basic families of 2/1 resonant periodic orbits, for the masses of the above system. Along each of these families the eccentricities of the two planets increase, starting with almost zero values, while the semimajor axes remain almost constant, corresponding to the 2/1 resonance. Large stable regions exist, even for large values of the two eccentricities. It is close to these stable periodic orbits that we should look for real planetary systems, because these are the only regions of phase space where bounded motion could exist. The exploration of the phase space close to a periodic orbit is made by studying perturbed orbits, by the method of Poincaré surface of section. It is shown that the phase (relative position of the two planets in their orbits) plays a crucial role on the stability of the system, and that for the same resonance and the same eccentricities chaotic motion appears, if the initial phase is changed, resulting to a quick disruption of the planetary system. On the other hand, a phase protection mechanism appears for suitable initial phase of the two planets, close to a stable periodic orbit, resulting to stable, ordered, motion.

## 2 Families of Periodic Orbits

### 2.1 Periodic Orbits in the Rotating Frame

As we mentioned above, the motion of an extrasolar planetary system is a special case of the general three body problem. We consider here the planar case. The best way to study this problem is to consider a rotating frame of reference (Hadjidemetriou 1975). This rotating frame is defined as follows:

Let us consider three bodies (point masses), $S$, $P_1$ and $P_2$, with masses $m_0$, $m_1$ and $m_2$, respectively, moving in the same plane under their mutual gravitational attraction. We define a rotating frame of reference, $xOy$, whose $x$ axis is the line $S\,P_1$, and the origin $O$ is the center of mass of these two bodies. The $y$ axis is perpendicular to the $x$ axis, in the plane of motion. The body $P_1$ moves on the $x$ axis and the body $P_2$ moves in the $xOy$ plane. This is a non uniformly rotating frame and the position of the three bodies are defined by the coordinate $x_1$ of $P_1$ on the $x$ axis, the coordinates $x_2$, $y_2$ of $P_2$ on the $xOy$ plane and the angle $\theta$ of the rotating $x$ axis with a fixed direction in the inertial frame (where the center of mass of the system of the three bodies is at rest). We have four degrees of freedom, but it turns out (Hadjidemetriou 1975) that the angle $\theta$ is ignorable, so we are left with three degrees of freedom in the rotating frame.

The initial conditions defining the motion of the three bodies in the rotating frame are: $x_{10}$, $x_{20}$, $y_{20}$, $\dot{x}_{10}$, $\dot{x}_{20}$, $\dot{y}_{20}$. Two more initial conditions are needed to completely define the motion in space: $\theta_0$ and $\dot{\theta}_0$. The value of $\dot{\theta}_0$ determines the angular momentum $L$, which appears as a fixed parameter in the differential equations of motion in the $xOy$ frame. The value of $\theta_0$ defines the initial orientation of the rotating frame and does not affect the motion. So, in order to study the motion in the rotating frame, we need the initial conditions $x_{10}$, $x_{20}$, $y_{20}$, $\dot{x}_{10}$, $\dot{x}_{20}$, $\dot{y}_{20}$ and the angular momentum $L$.

It can be proved that *periodic orbits* of the three body system exist in the rotating frame defined above (Hadjidemetriou 1975). Note that periodicity on the rotating frame means that the *relative* configuration of the three bodies is repeated after one period. The system *is not*, in general, periodic in the inertial frame.

It is proved (Hadjidemetriou 1976) that the periodic orbits in the rotating frame are not isolated, but belong to *monoparametric families of periodic orbits*, along which all the masses are fixed.

Of particular interest are the *symmetric periodic orbits*. The numerical integrations that follow showed that all the basic families of periodic orbits turned out to be symmetric with respect to the $x$ axis of the rotating frame. A periodic orbit is symmetric with respect to the $x$ axis of the rotating frame $xOy$ if at $t = 0$ the body $P_1$, which moves always on the $x$ axis, has zero velocity ($\dot{x}_{10} = 0$) and the body $P_2$ is on the $x$ axis and its velocity is perpendicular to this axis ($y_{20} = 0$, $\dot{x}_{20} = 0$). Consequently, a symmetric periodic orbit is determined from three nonzero initial conditions only, namely $x_{10}$, $x_{20}$ and $\dot{y}_{20}$. So, such a family can be represented by a continuous curve in the space $x_1, x_2, \dot{y}_2$.

### 2.2 Periodic Orbits of the Planetary Type

In section 2.1 we mentioned that the general planar three body problem can be reduced to a system of three degrees of freedom, in a rotating frame $xOy$. We shall now restrict ourselves to a planetary system with the sun $S$, and two planets $P_1$ and $P_2$, with the mass $m_0$ of the sun much larger than the masses $m_1$ and $m_2$ of the planets. This is a special case of the general three body problem, but now, since the masses $m_1$ and $m_2$ are small, we can consider this system as a *perturbed* system of the *integrable system* consisting of two uncoupled two body systems, $S - P_1$ and $S - P_2$.

Let us assume at first that $m_1 = 0$ and $m_2 = 0$ and that the planets $P_1$ and $P_2$ move in *circular* orbits, with radii $a_1$ and $a_2$, respectively, in the same plane and the same direction. It is clear that this motion is a symmetric periodic motion, with respect to the $x$ axis, in the rotating frame $xOy$ defined above, for any value of $a_1$ and $a_2$, i.e. for any value of the ratio of the periods of revolution of the two planets around the sun $S$. So, we have a monoparametric family of nonresonant, in general, periodic orbits in the unperturbed problem (zero planetary masses).

Next, we consider the case where the masses of the planets are zero and they move around the sun in *elliptic* orbits, but their semimajor axes are such that the ratio $T_1/T_2$ of their periods is a rational number. This means that we are in a *mean motion resonance*. We keep now the semimajor axes of the planets fixed and vary the eccentricities and the orientation of the two uncoupled planetary orbits. The resulting motion is periodic (not symmetric, in general) in the rotating $xOy$ frame. So, we have families of unperturbed resonant periodic orbits in the rotating frame.

We give now to the planets nonzero masses. What happens to the above mentioned unperturbed families? It can be proved (Hadjidemetriou 1976) that all the members of the unperturbed family of circular orbits can be continued to the nonzero case, as *symmetric circular* periodic orbits. This is true for all orbits, except for those corresponding to the resonances of the form $(n+1)/n$, $n = 1, 2, 3, ...$, i.e. to the resonances 2/1, 3/2, .. At these points gaps appear and the single unperturbed family of circular periodic orbits breaks into an infinite number of families of symmetric nearly circular periodic orbits, which are separated at the resonances 2/1, 3/2, ...by gaps. At these gaps, a bifurcation of a family of *resonant elliptic* symmetric periodic orbits appears and the circular family continues as a resonant elliptic family of symmetric periodic orbits, along which the resonance remains constant, equal to the corresponding resonance at the bifurcation point, and the eccentricities of the two planets increase, starting from almost zero values, as we go outwards. We remark at this point that these families of elliptic periodic orbits are the continuation (from zero to non zero planetary masses) of the unperturbed families of elliptic periodic orbits that we mentioned above.

So, in the space $x_{10}$, $x_{20}$, $\dot{y}_{20}$ of nonzero initial conditions, the perturbed families of symmetric periodic orbits are represented by a set of continuous curves, having a circular, nonresonant in general, part and a resonant part, with a fixed resonance. This is shown in Fig. 1, for the region near the 2/1 and 3/2

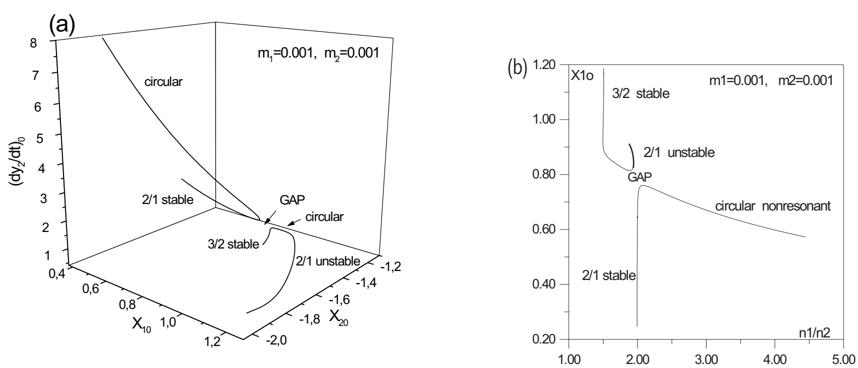



**Fig. 1.** Families of circular and elliptic periodic orbits of the general three body problem, of the planetary type, for $m_1 = 0.001$, $m_2 = 0.001$ and $m_{sun} = 0.998$. A discontinuity exists at the 2/1 resonance and branches of 2/1 resonant elliptic periodic orbits appear. Also, a stable 3/2 resonant branch of periodic orbits appears: (a) The families in the $x_{10}$, $x_{20}$, $\dot{y}_{20}$ space. (b) The families in the *resonance* - $x_0$ space. One 2/1 resonant branch (marked by a thick line) is unstable. There is also a small region on the circular branch close to the 3/1 resonance (not shown) which is also unstable.

resonances, for the masses $m_0 = 0.998$, $m_1 = 0.001$ and $m_2 = 0.001$. Note the gap at the 2/1 resonance.

The orbits of the two planets of a periodic system in the *inertial* frame are nearly Keplerian ellipses, due to their week gravitational interaction. These ellipses precess with a small angular velocity, which depends on the particular periodic orbit, if the motion is ordered (see Fig. 3b, as an example), but the precession may be chaotic in some cases.

So, summarizing all the above, we can say that there are two types of periodic orbits of the planetary type, *circular* and *resonant elliptic*:

- *Circular*: The orbits of the two planets are almost circular, and
- *resonant elliptic*: The orbits of the two planets have finite eccentricities, but the two semimajor axes are such that the ratio of the periods of the two planetary orbits is rational.

The basic periodic orbits are *symmetric* with respect to the rotating $x$-axis. This means that at $t = 0$ the initial conditions are $x_{10}$, $\dot{x}_{10} = 0$, $x_{20}$, $y_{20} = 0$, $\dot{x}_{20} = 0$, $\dot{y}_{20}$. In the study of the motion in the rotating frame, the angular momentum is kept as a fixed parameter and can be used to find $\dot{\theta}$.

### 2.3 Families of Periodic Orbits at the 2/1 Resonance

We make now a complete study of the *periodic orbits*, close to the 2/1 resonance, for a planetary system having the masses of the extrasolar planetary system HD82943, namely

$$m_1 = 0.88J, \;\; m_2 \sin i = 1.63J, \;\; m_{sun} = 1.05M_\odot, \tag{1}$$

(for $\sin(i) = 1$). The mass of the sun is in solar masses and the masses of the two planets are in Jupiter masses.

In order to avoid duplication in the numerical study, we must fix the units of mass, length and time. In the present study this normalization is made by taking the total mass of the system equal to one, $m_0 + m_1 + m_2 = 1$, the gravitational constant equal to one and we keep the angular momentum $L$ fixed, equal to $L = 0.002$. This means that the unit of mass is slightly larger than one solar mass. Concerning the units of length and time, we note that they depend on the particular value of the angular momentum $L$ that we are using in the numerical computations (we remind that $L$ appears as a fixed parameter in the equations of motion in the rotating frame). So, a particular periodic motion of the planetary system may correspond to an infinite set of periodic planetary systems with *similar* planetary orbits. There are however some elements of the orbit whose values are independent of the units, namely the eccentricities of the planets or the ratio of the planetary semimajor axes (or planetary periods). Several plots that follow are in these elements of the orbit.

In normalized units, the masses of the planetary system that we will study are

$$m_1 = 0.0008, \;\; m_2 = 0.0014, \;\; m_0 = 0.9978. \tag{2}$$

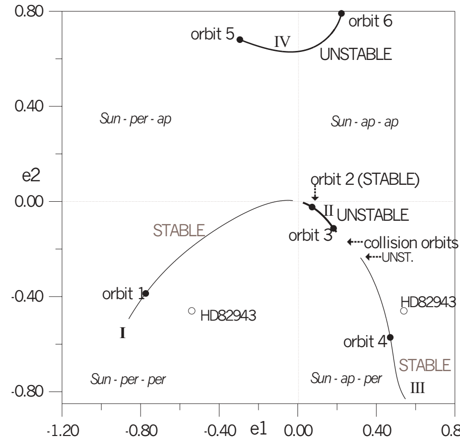



**Fig. 2.** The families *I, II, III, IV* of periodic orbits for the masses of HD82943, in the eccentricity space. The indication *per, ap* refers to the initial configuration at $t = 0$. (*per* denotes position at periastron and *ap* position at apoastron.)

We computed four different families of symmetric periodic orbits, all at the 2/1 resonance, which differ in the initial phase of the two planets at an initial epoch $t = 0$. These families are shown in Fig. 2. Along each family the angular momentum has a fixed value, $L = 0.002$, the same for all members of the family. To make the presentation clearer, we present these families in the space of the initial eccentricities of the two planets. These are in fact the osculating eccentricities at $t = 0$, but we note that they do not vary much along the periodic orbit, unless we are close to a collision orbit. At $t = 0$ the two planets and the sun are in the same line, the planets being either at periastron or apoastron. This is a consequence of the fact that the periodic orbits are symmetric with respect to the $x$ axis of the rotating frame. Throughout this study we will assume that the planet $P_1$ is the inner planet and the planet $P_2$ is the outer planet.

In Fig. 2 we made the convention to use a positive value of the eccentricity if the corresponding planet is at apoastron at $t = 0$ and a negative eccentricity if it is at periastron. Note that due to the fact that we are in a 2/1 resonance, a simple geometric consideration shows that the outer planet $P_2$ after a half period $t = T/2$ changes position from periastron to apoastron, or vice versa, while the inner planet $P_1$ returns to the same position as at $t = 0$.

We remark that the families designated as $II$ and $III$ in Fig. 2 are in fact a single family, and the gap that appears between them is simply due to the existence of a collision orbit between the two planets, so that the computation presented numerical difficulties.

The whole family $I$ is stable. Family $II$ is mainly unstable, with the exception of a small region which is stable. Family $III$ is mainly stable, and there is an unstable region close to the collision orbit. Finally, the family $IV$ is unstable.

In each periodic planetary system we have two nearly Keplerian planetary orbits, which are not fixed in space, due to the gravitational interaction between the two planets. The line of apsides of both planets rotates slowly.

## 3 The Different Types of 2/1 Resonant Periodic Orbits

Some typical orbits of each of the above four families are presented in the Figs. 3–7. We find it more illustrating to present the orbits in the inertial frame, although, as we mentioned, there is a slow precession of the line of apsides, because these orbits are periodic in a rotating frame only, in general. The orbits thus shown are for a small time period, close to the period of the periodic orbit in the rotating frame. In addition, we also present some periodic orbits in the rotating frame, to obtain an idea how the motion looks in the synodic frame.

In all the periodic orbits, the line of apsides of the two planetary orbits coincide. The pericenters of the two orbits may be in the same direction, with respect to the sun, or in opposite directions. Concerning the phase, i.e. the position of the two planets in their orbits at $t = 0$, all possible combinations can appear. Due however to the 2/1 resonance between the two planets, if we start with a certain phase at $t = 0$, after half a period, $t = T/2$, the position of $P_2$ shifts from pericenter to apocenter, or vice versa. So we have, for each periodic motion, two equivalent configurations, at $t = 0$ and at $t = T/2$, respectively. To the above two different phases of the same orbit, we have two perpendicular crossings of the planet $P_2$ from the rotating $x$ axis, and consequently we could use either perpendicular crossing to represent a periodic orbit in Fig. 2.

**Table 1.** All possible phases at $t = 0$ and $t = T/2$

| | | | |
|---|---|---|---|
| *Type 1*: | Sun - $P_1$(per) - $P_2$(per) | $\rightarrow$ | $P_2$(ap) - Sun - $P_1$(per) |
| *Type 2*: | Sun - $P_1$(ap) - $P_2$(ap) | $\rightarrow$ | $P_2$(per) - Sun - $P_1$(ap) |
| *Type 3*: | Sun - $P_1$(per) - $P_2$(ap) | $\rightarrow$ | $P_2$(per) - Sun - $P_1$(per) |
| *Type 4*: | Sun - $P_1$(ap) - $P_2$(per) | $\rightarrow$ | $P_2$(ap) - Sun - $P_1$(ap) |

All the possible initial phases of a periodic orbit, and the equivalent configuration at $t = T/2$, are summarized in Table 1. As we will see in the next subsections, the 2/1 resonant orbits with a phase of type 1 are stable, because a phase protection mechanism appears. These are the orbits of the family $I$. The orbits of the families $II$ and $III$ are of the type 4. These orbits are also stable, provided that we are not close to a collision orbit, and the eccentricities of the two planets are large. In contrast, the orbits of the type 4 are unstable when the eccentricities of the two planets are small, as is the case with the family $II$ (with some exceptions), and when we are close to a collision orbit. The resonant orbits of the type 2 and 3 belong to the family $IV$ and are all unstable.

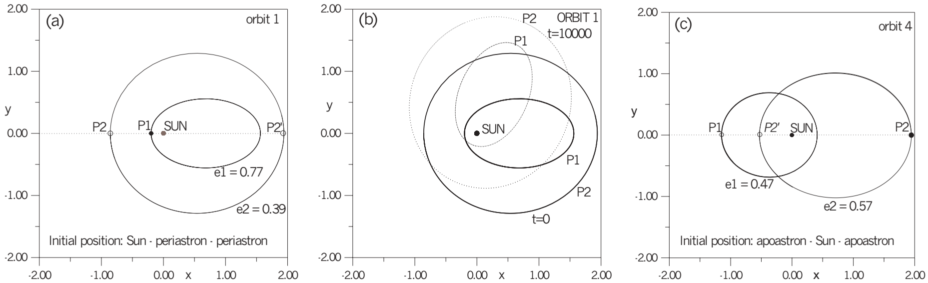



**Fig. 3.** (a) The *orbit 1* of Fig. 2 for a short time. (b) The same orbit at two different time intervals. The precession of the line of apsides is clear. (c) The *orbit 4* of Fig. 2. Note that both orbits 1 and 4 have large planetary eccentricities, but the system is stable, even for the *orbit 4*, where the orbits of the planets intersect.

### 3.1 Stable Orbits of the Families *I* and *IV*

A typical orbit of the family *I* is shown in Fig. 3a. The exact position of this orbit on the family *I* of Fig. 2 is indicated as *orbit 1*. The phase of this orbit is of type 1 of the Table 1. This motion is stable, although the eccentricities are large. The two planetary orbits do not intersect in space.

A typical orbit of the family *III* is shown in Fig. 3c. The exact position of this orbit on the family *III* of Fig. 2 is indicated as *orbit 4*. This orbit is also stable, with large values of the eccentricities, but now, in contrast to the orbits of the family *I*, the planetary orbits intersect. The initial phase is of type 4 in Table 1, and due to the fact that the two planets are locked to the 2/1 resonance, a phase protection mechanism exists, which prevents the two planets from close encounters, although their orbits intersect in space. A 2/1 resonant stable orbit close to the system HD82943, of the type of the Fig. 3c, where the two planetary orbits intersect, was presented last September 2002 by Ji *et.al* at the IAU Colloquium 189.

### 3.2 A Stable and an Unstable Orbit of the Family *II*

We present in Fig. 4 two orbits of the family *II*, one stable (Fig. 4a) and one unstable (Fig. 4b). The exact position of these orbits on the family *II* of Fig. 2 is indicated as *orbit 2* and *orbit 3*, respectively. Both orbits are of type 4 of the Table 1, as was the case with the orbit 4 of the Fig. 3b. In fact, the families *II* and *III* belong to the same family, as already mentioned before, but they appear as separated in Fig. 2 because there exists a collision periodic orbit between the two planets, and the gap which appears close to this collision orbit is due to numerical difficulties. All the orbits close to the collision orbit are strongly unstable.

We note that the eccentricities of the two planets are rather small, but now the *orbit 3* (Fig. 4b) is unstable. The *orbit 2* (Fig. 4a) has smaller eccentricities and is stable.

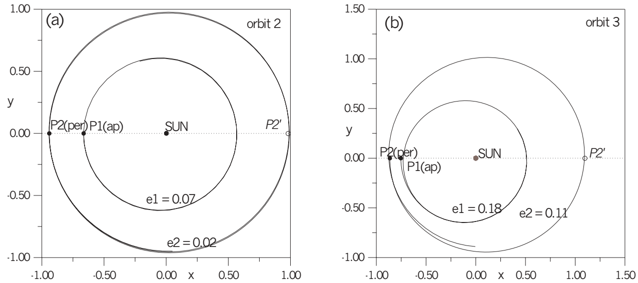



**Fig. 4.** (a) The *orbit 2* and (b) the *orbit 3* of Fig. 2. Note that in both cases the orbits of the planets have small eccentricities and the phase is the same, but the *orbit 2* is stable while the *orbit 3* is unstable.

It is interesting to compare the *orbit 3* (Fig. 4) with the *orbit 4* (Fig. 3). They have the same phase, but the *orbit 3* is unstable, while the *orbit 4* is stable. The instability of the *orbit 3* is due to the close approach between the two planets, as is clearly seen from the plot of Fig. 4. In the *orbit 4* the closest approach between the two planets is much larger, due to the large value of the planetary eccentricities. Thus, the increase of the eccentricities plays a stabilizing role, because, in relation also to a phase protection mechanism due to the fact that the two planets are locked to the 2/1 resonance, its effect is to increase the closest approach between the planets. This may explain why there are several actual extrasolar planetary systems with large eccentricities.

### 3.3 Two Orbits of the Unstable Family *IV*

We present in Fig. 5 two typical periodic orbits of the family $IV$, indicated as *orbit 5* (Fig. 5a) and *orbit 6* (Fig. 5b) of Fig. 2, at the two ends of the curve representing this family. The orbit 5 has a phase of type 2 in Table 2, while the orbit 6 has a phase of type 3. Note that the phase changes along the family from type 2 to type 3. This is due to the fact that the eccentricity of the inner planet starts with a positive value (*orbit 5*, position at apoastron) and as we proceed

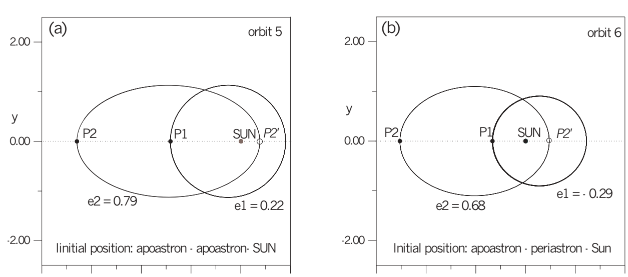



**Fig. 5.** (a) The *orbit 5* and (b) the *orbit 6* of Fig. 2. Note that in both cases the orbits of the planets have large eccentricities and, contrary to the *orbit 4*, both systems are unstable.

along the family it changes to a negative value (*orbit 6*, position at periastron), passing through a circular orbit, as is clearly seen in Fig. 2. Along the whole family *IV* the eccentricity of the outer planet is large. The orbits of the planets intersect in space, as in the family *III* (*orbit 4*), but now the phase is such that the phase protection mechanism is not operating and both orbits are unstable.

### 3.4 The Periodic Orbits in the Synodic Frame

In Fig. 6 we present the periodic orbits 1 and 4 of the Figs. 3a and 3c, respectively, in the rotating $xOy$ frame, in which they are exactly periodic.

In Fig. 7 we present the periodic orbits 5 and 6 of the Figs. 5a and 5b, respectively, in the rotating $xOy$ frame. Note that in the rotating frame the planet $P_1$ moves on the $x$ axis only and its orbit is presented as a straight line. The length of this line depends on the eccentricity.

## 4 Perturbed Orbits Close to Stable and Unstable Periodic Orbits

In the previous sections we computed four basic families of resonant periodic orbits, at the 2/1 resonance. We shall explore now the phase space close to these periodic orbits, in order to detect the regions where stable motion exists.

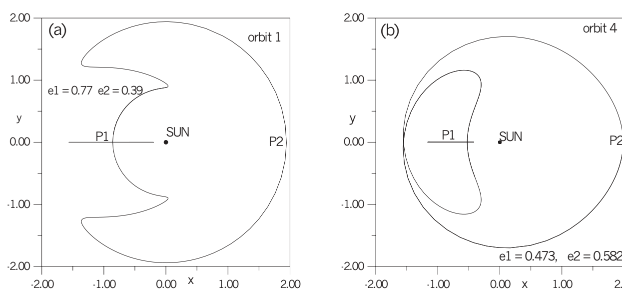



**Fig. 6.** The *orbit 1* of Fig. 2 and the *orbit 4* of Fig. 2, in the rotating frame.

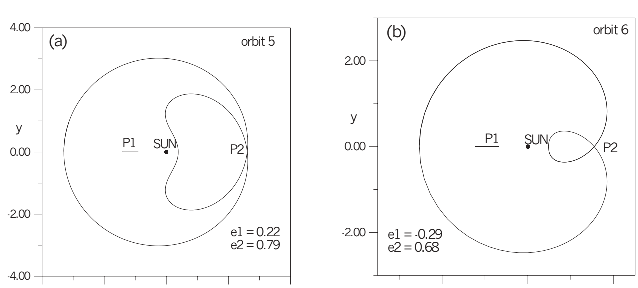



**Fig. 7.** The *orbit 5* of Fig. 2 and the *orbit 6* of Fig. 2, in the rotating frame.

We shall also study the generation of chaotic motion and the disruption of the planetary system.

The dynamical system we study has three degrees of freedom, in the rotating frame, so its phase space is six-dimensional. This can be reduced to a four-dimensional phase space, if we take the Poincaré map which is on a four-dimensional surface of section. A complete search of the whole phase space is very tedious, even if we restrict ourselves to the vicinity of a periodic orbit (a fixed point on the Poincaré map). In the following we study the effect of a phase shift of the planets, *in their orbits*, on the stability of the system. We start with a periodic orbit belonging to one of the four families presented in Fig. 2, and we change the initial position of the planet $P_2$, *on its orbit.* In his way we obtain a new planetary system, which is not periodic, but the two planetary orbits have the same elements. The study of the long term evolution of the perturbed system is made by computing the Poincaré map. We have taken as surface of section the surface $y_2 = 0$ and *Energy*=constant. So the phase space of the mapping is the four-dimensional space $x_1$, $\dot{x}_1$, $x_2$, $\dot{x}_2$ (the coordinates are in the rotating frame $xOy$).

In the following we show, for each orbit we studied, the projection of the mapping on one of the coordinate planes and also the time evolution of the eccentricities and the semimajor axes.

### 4.1 Perturbed Orbits Close to the *Orbit 1* and the *Orbit 4*

We start with two stable periodic orbits, *orbit 1* and *orbit 4*, belonging to the families $I$ and $III$, respectively. Our aim is to see how the motion is affected if the planet $P_2$ is shifted along its orbit, thus changing the phase of the system. In Figs. 8a and 8b we present two shifted positions of $P_2$ along its orbit, indicated as $P_{21}$ and $P_{22}$, for the periodic orbit 1 (Fig. 3a) and the periodic orbit 4 (Fig. 3b), respectively. The results are given in Figs. 9 and 10 for the perturbed orbits to the orbit 1 and in Figs. 11, 12 for the perturbed orbits to the orbit 4.

Note that for a small shift of $P_2$ on its orbit, to the position $P_{21}$, the perturbed motion is on a torus, as is clearly seen from the Poincaré maps in Figs. 9a and 10a. We remark that in these Figures we presented the projection of the

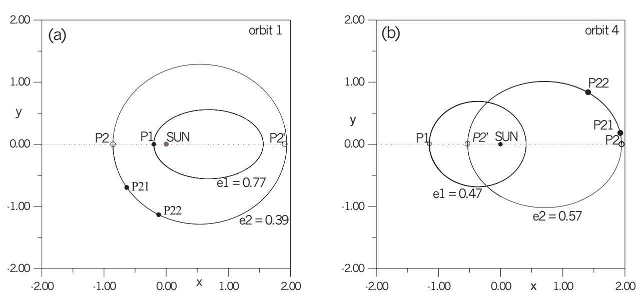



**Fig. 8.** The shift of the planet $P_2$ along its orbit, to the positions $P_{21}$ and $P_{22}$, on the *orbit 1* and on the *orbit 4*.

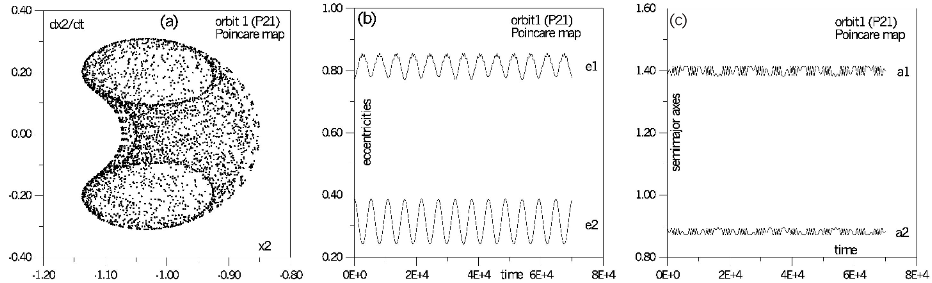



**Fig. 9.** A perturbed orbit of the *orbit 1*, due to a shift of $P_2$ along its orbit to the position $P_{21}$ of Fig. 8a: (a) Projection of the four dimensional Poincaré map in the space $x_2\dot{x}_2$, (b) the evolution of the eccentricities in time, (c) the evolution of the semimajor axes in time. The perturbed motion is on a torus.

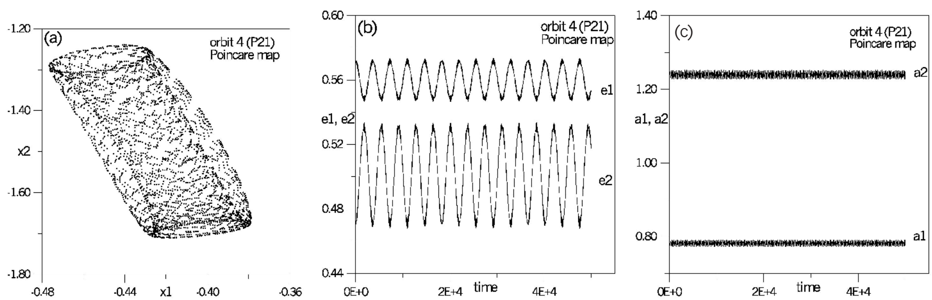



**Fig. 10.** A perturbed orbit of the *orbit 4*, due to a shift of $P_2$ along its orbit to the position $P_{21}$ of Fig. 8b: (a) Projection of the four dimensional Poincaré map in the space $x_1x_2$, (b) the evolution of the eccentricities in time, (c) the evolution of the semimajor axes in time. The perturbed motion is on a torus.

four dimensional Poincaré map on a two dimensional coordinate plane. The projection of the Poinaré map in all the other coordinate planes (not shown here) is similar. So, indeed the perturbed motion on the surface of section is on a four dimensional torus. The evolution of the semimajor axes and the eccentricities of both planets undergo oscillations with a fixed amplitude. This is an indication that in both cases the motion is ordered and the two planets move on bounded orbits. This means that a small change of the phase of a stable periodic orbit results to ordered, bounded, motion on a torus in the four dimensional phase space.

If the shift of $P_2$ along its orbit is larger, to the position $P_{22}$, resulting to a larger change of the phase of the system, the stable periodic orbits 1 and 4 become now unstable and the system quickly disrupts into a close binary, consisting of the sun and one planet, while the other planet escapes to infinity. This is clearly seen in Figs. 11 and 12, for the orbits 1 and 4, respectively.

From the above we see that a small shift of the phase close to a periodic motion results to ordered motion on a torus, while a larger shift of the phase results to instability and escape of one planet. The distinction between ordered

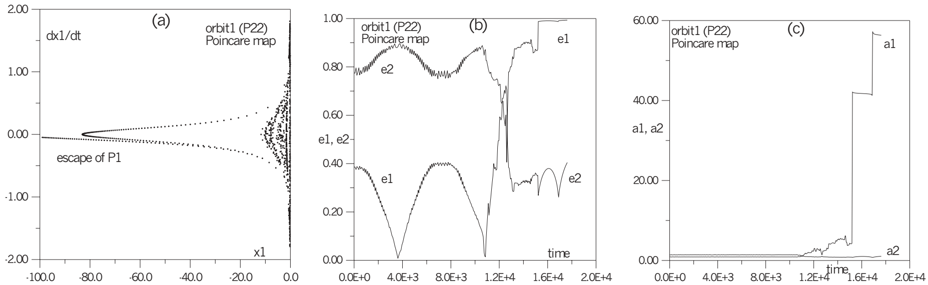



**Fig. 11.** A perturbed orbit of the orbit 1, due to a shift of $P_2$ along its orbit to the position $P_{22}$ of Fig. 8a: (a) Projection of the four dimensional Poincaré map in the space $x_1\dot{x}_1$, (b) the evolution of the eccentricities in time, (c) the evolution of the semimajor axes in time. The perturbed motion starts on a perturbed torus, but soon the planet $P_1$ escapes.

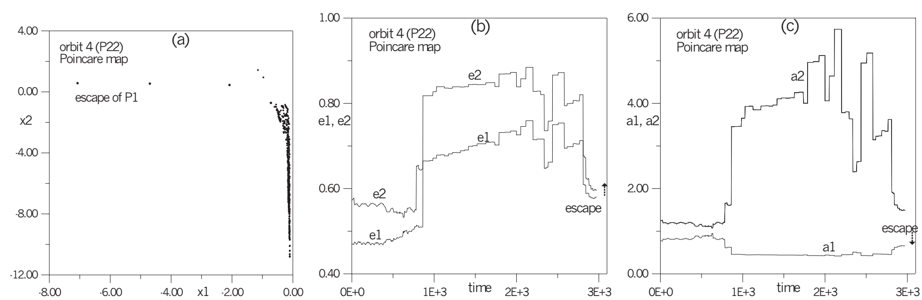



**Fig. 12.** A perturbed orbit of the orbit 4, due to a shift of $P_2$ along its orbit to the position $P_{22}$ of Fig. 8b: (a) Projection of the four dimensional Poincaré map in the space $x_1x_2$, (b) the evolution of the eccentricities in time, (c) the evolution of the semimajor axes in time. The perturbed motion starts on a perturbed torus, but soon the planet $P_1$ escapes.

and chaotic motion was based on the geometric properties of the Poincaré map, as obtained from the numerical computations. A more rigorous analysis, based on spectral analysis of the time series of several elements of the orbit (eccentricity, semi major axis) verified the distinction between order and chaos, as obtained from the graphs. An typical example of ordered and of chaotic motion is presented in the next section.

## 4.2 Perturbed Orbits Close to the *Orbit 2* and the *Orbit 3*

As in the previous subsection, we study the effect of a phase shift of the planet $P_2$ on its orbit. We consider two orbits of the family *II*, *orbit 2* and *orbit 3*. The first is stable and the second is unstable. The shifted positions are shown in Fig. 13. The results of the numerical computations are shown in Figs. 14 and 15 for the *orbit 2* and in Figs. 15–18 for the *orbit 3*.

The numerical results show that a shift of $P_2$ along the stable periodic *orbit 2* results to ordered, bounded, motion on a torus, even for a large change of

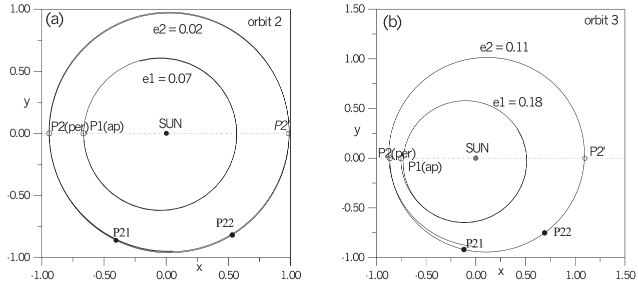



**Fig. 13.** The shift of the planet $P_2$ along its orbit, to the positions $P_{21}$ and $P_{22}$ on the *orbit 2* and on the *orbit 3*.

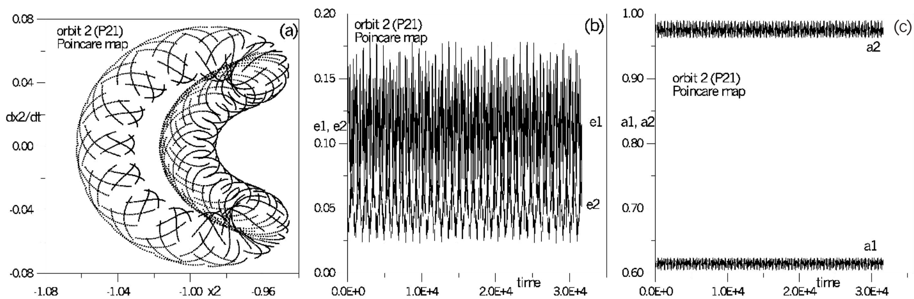



**Fig. 14.** A perturbed orbit of the *orbit 2*, due to a shift of $P_2$ along its orbit to the position $P_{21}$ of Fig. 13a: (a) Projection of the four dimensional Poincaré map in the space $x_2\dot{x}_2$, (b) the evolution of the eccentricities in time, (c) the evolution of the semimajor axes in time. The perturbed motion is on a torus.

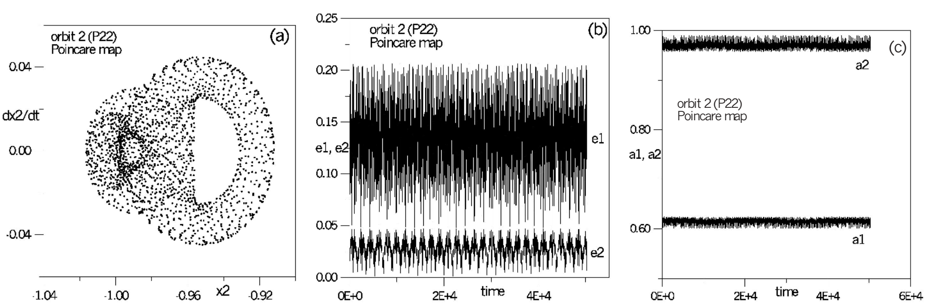



**Fig. 15.** A perturbed orbit of the *orbit 2*, due to a shift of $P_2$ along its orbit to the position $P_{22}$ of Fig. 13a: (a) Projection of the four dimensional Poincaré map in the space $x_2\dot{x}_2$, (b) the evolution of the eccentricities in time, (c) the evolution of the semimajor axes in time. The perturbed motion is on a torus.

the phase (positions $P_{21}$ and $P_{22}$ of Fig. 13a), as shown clearly in the Figs. 14 and 15. (We also mention here that the projection of the Poincaré map to all other coordinate planes was similar to that shown in Figs. 15 and 15).

In contrast, a shift of $P_2$ along the unstable periodic orbit 3 (Fig. 13b) results to chaotic motion, although no escape was detected. Note that when the phase

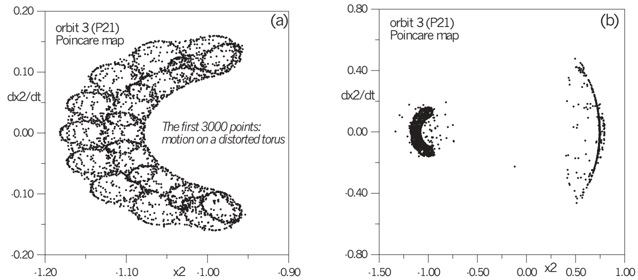



**Fig. 16.** A perturbed orbit of the *orbit 3*, due to a shift of $P_2$ along its orbit to the position $P_{21}$ of Fig. 13b: Projection of the four dimensional Poincaré map in the space $x_2\dot{x}_2$ (a) for the first 3000 iterations, and (b) for long time. The motion starts on a torus, but later it becomes chaotic.

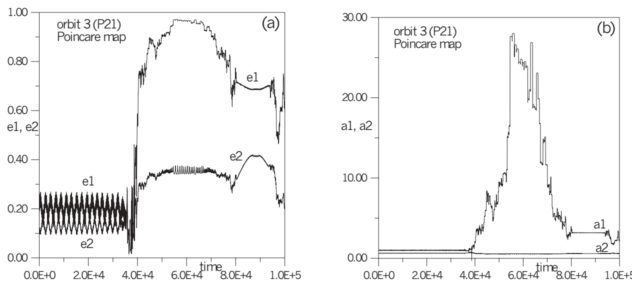



**Fig. 17.** A perturbed orbit of the *orbit 3*, due to a shift of $P_2$ along its orbit to the position $P_{21}$ of Fig. 13b: (a) Poincaré map for the evolution of the eccentricities in time and (b) the evolution of the semimajor axes in time.

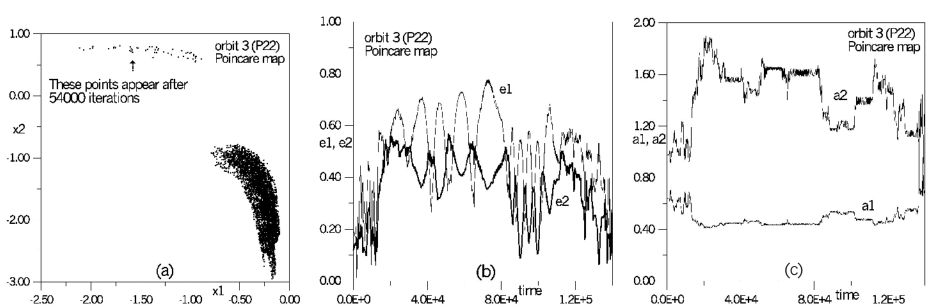



**Fig. 18.** A perturbed orbit of the *orbit 3*, due to a shift of $P_2$ along its orbit to the position $P_{22}$ of Fig. 13b: (a) Projection of the four dimensional Poincaré map in the space $x_1x_2$, (b) the evolution of the eccentricities in time, (c) the evolution of the semimajor axes in time. The perturbed motion is chaotic.

shift is small (position $P_{21}$), the perturbed motion starts on a distorted torus (Fig. 16a), but later on the motion becomes chaotic (Figs. 16b, 17). Note that both the eccentricities and the semimajor axes are bounded in an oscillatory motion for a long time initially, corresponding to the motion on the distorted torus in Fig. 16a, but after a certain time chaotic motion starts. This phenomenon is in fact a manifestation of the *stickiness effect*, studied in Efthimiopoulos et. al. (1997) for two degrees of freedom.

For a larger phase shift of $P_2$, to the position $P_{22}$ in Fig. 13b, the perturbed motion is chaotic from the start (Fig. 18).

### 4.3 Perturbed Orbits Close to the *Orbit 5*

In Fig. 19 we show the shift of the planet $P_2$ to the positions $P_{21}$ and $P_{22}$, along the unstable periodic *orbit 5* of the family $IV$. The resulting motion is unstable for all positions of $P_2$. An example is given in Fig. 20, for a shift to the position $P_{21}$, where, after a strong chaotic motion of the system, the planet $P_1$ escapes. The same evolution appears for the shift to the position $P_{22}$.

## 5 Discussion

We have presented a systematic way to find all the regions of the phase space of a planetary system with two planets, moving in the same plane, where stable

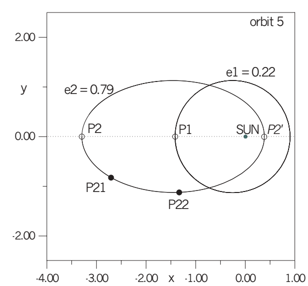



**Fig. 19.** The shift of the planet $P_2$ along its orbit, to the positions $P_{21}$ and $P_{22}$, on the *orbit 5.*

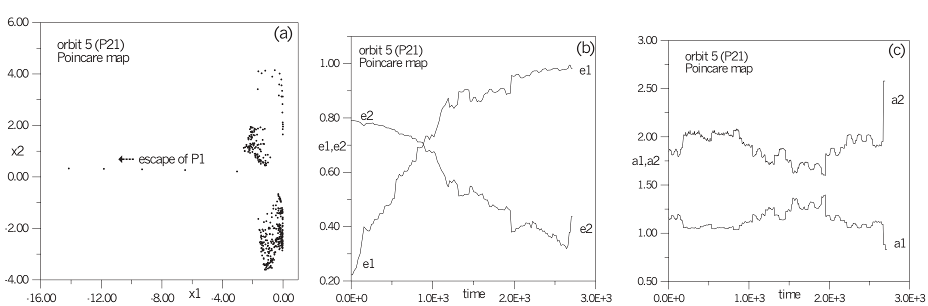



**Fig. 20.** A perturbed orbit of the *orbit 5*, due to a shift of $P_2$ along its orbit to the position $P_{21}$ of Fig. 19: (a) Projection of the four dimensional Poincaré map in the space $x_1x_2$, (b) the evolution of the eccentricities in time, (c) the evolution of the semimajor axes in time. The planet $P_1$ escapes. In (b) and (c) the last points, leading to escape, are not shown.

motion could appear. This was achieved by studying the families of periodic orbits and their stability properties. The study was restricted to the region close to the 2/1 mean motion resonance between the planetary orbits, but a similar study can be carried out for all other important resonances. Each resonance however has its own characteristics, and the behaviour is different from one type of resonance to the other. The present study is restricted to planetary systems having the same masses as the observed planetary system HD82943. But we note at this point that a change of the planetary masses, for example keeping the total mass fixed and varying the ratio $m_1/m_2$, may result to a change of the stability (Hadjidemetriou, 2002).

The phase of the planets plays a crucial role on the stability of the system. This is so because, for some phases a *phase protection mechanism* is operating, protecting the two planets from close encounters. It was found that the system is stable, even for large eccentricities, provided that both planets start from their perihelia, situated in the same direction (*type 1* of Table 1). In this case the orbits of the planets do not intersect. This is the phase of all the orbits of the family $I$. Another stable phase, where the planetary orbits do intersect in space, is when the inner planet starts at aphelion and the outer planet at perihelion, both situated in the same direction (*type 4* of Table 1). It is interesting to note that for this phase, the increase of the eccentricities plays a stabilizing role. This is clear in Fig. 2, where the families $II$ and $III$ have the same phase (*type 4*), but almost all orbits of the family $II$, corresponding to small eccentricities, are unstable, while all orbits of the family $III$ (with the exception of those close to the collision orbit) are stable. A typical unstable orbit is shown in Fig. 4b and a typical stable orbit is shown in Fig. 3a (both for the same phase of *type 4*).

Contrary to these stable phases, there are two more phases that are in all cases unstable (*type 2* and *type 3* of Table 1). These are the phases of the orbits of the family $IV$. We did not find any stable periodic motion corresponding to these phases. For the orbits of the family $IV$ we found that a small deviation from the exact periodic motion results to a distorted motion and sooner or later a close encounter between the two planets appears, resulting to chaotic motion and in some cases disruption of the system (Fig. 20).

For the stable orbits, we studied the long term behaviour of the perturbed motion, obtained by shifting the planet $P_2$ along its orbit, thus changing the initial phase, but leaving all planetary elements unchanged. We found that for a small shift stable, quasi periodic, motion appears as is clear from the corresponding Poincaré maps that we have computed. Thus we found that there do exist zones of stability, where stable planetary systems could appear in nature. Some typical examples are in Figs. 9a, 10a, 14a and 15a. From the above it is clear that planetary systems with large eccentricities could be common on the sky. However, such systems with large eccentricities should necessarily be close to a resonance, with the proper phase, because it is the resonance that generates the phase protection mechanism which stabilizes the system. Far from the resonance close encounters are inevitable, resulting to chaotic motion.

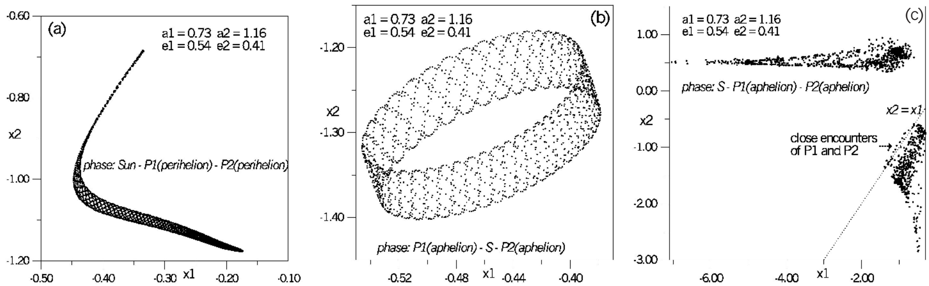



**Fig. 21.** The long run of a planetary orbit with the elements of HD82943 (Projection of the four dimensional Poincaré map in the space $x_1x_2$): (a) for the configuration Sun-$P_1$(perihelion)-$P_2$(perihelion), (b) for the configuration $P_1$(aphelion)-Sun-$P_2$(aphelion) and (c) for the configuration Sun-$P_1$(aphelion)-$P_2$(aphelion). The repeated close encounters in (c) are clearly seen, close to the line $x_2 = x_1$ (note that $y_1 = 0$, $y_2 = 0$).

The observed planetary system HD82943 is close to the stable configurations we found in our study, as shown in Fig. 2. In Fig. 21 we present the Poincaré map using the orbital elements of the system HD82943. Since the phase is not known, we used three different phases, one of the stable *type 1* of Table 1, corresponding to the family *I* (Fig. 21a), one of the stable *type 4*, corresponding to the family *III* (Fig. 21b) and one of the unstable *type 2*, corresponding to the family *IV* (Fig. 21c). We note that in the two stable phases the system is bounded, moving on a torus, while the motion for the unstable phase is chaotic, resulting to large variations of the planetary orbits and eventually to escape of one planet. Note that in this latter case close encounters between the two planets appear, as indicated in Fig. 21c, resulting to strong chaotic motion and large changes of the orbital elements.

The ordered or the chaotic nature of the motion in Fig. 21, as is clearly seen on the projections of the Poincaré map, is verified by a power spectrum analysis of the time series of the elements of the orbit, along the perturbed motion. In Fig. 22 we present two typical power spectra, one for the ordered motion of the Fig. 21b and one for the chaotic motion of the Fig. 21c, for the time series for the

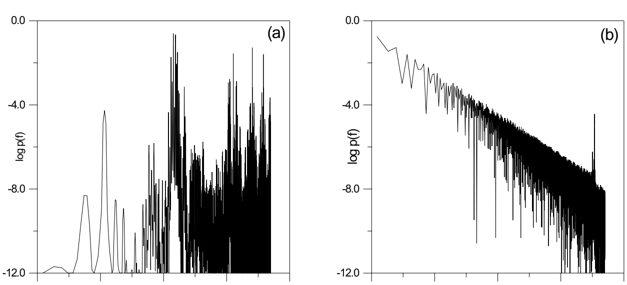



**Fig. 22.** Power spectra of the time series of the evolution of the semimajor axis: (a) for the ordered orbit of the Fig. 21b and (b) for the chaotic orbit of the Fig. 21c.

semimajor axis $a_1$. The distinction between order and chaos is clear. In Fig. 22a the power spectrum density has the typical features of ordered motion, (a few peaks of significant amplitude). On the other hand, the power spectrum density in Fig. 22b is typical of a chaotic motion, because it shows $1/f$-divergence, which indicates strong chaos.

A similar power spectrum density analysis was made for all the orbits that we studied in the previous sections and the geometric picture of order or chaos that we observed on the Poincaré map was confirmed.

**Acknowledgements**

We thank Dr. G. Voyatzis for his help in the the power spectrum analysis. We also thank the referee for his useful comments.

## References

24. Beauge C., Ferraz-Mello S. & Michtchenko T. (2002): *Extrasolar Planets in Mean-Motion Resonance: Apses Alignment and Asymmetric Stationary Solutions*, preprint.
24. Callegari N., Michtchenko T. & Ferraz-Mello S. (2002): *Dynamics of two planets in the 2:1 and 3:2 mean motion resonances*, in 34th DPS Meeting, BAAS, 34, 30.01
24. Celletti A., Chessa A., Hadjidemetriou J.D. and Valsecchi J.B. (2002), *A systematic study of the stability of symmetric periodic orbits in the planar, circular, restricted three-body problem*, Cel. Mech. & Dyn. Astron. **83**, 239-255.
24. Efthimiopoulos C., Contopoulos G., Voglis N., and Dvorak R. (1997): *Stickiness and cantori*, Journal of Physics A - Mathematical and General **30**, 8167-8186.
24. Ferraz-Mello S. (2002): *Tidal Acceleration, Rotation and Apses Allignement in Resonant Extra-Solar Planetary Systems*, in 34th DPS Meeting, BAAS, 34, 30.08
24. Ford E.B, Havlikova M. and Rasio F.A. (2001): Icarus **150**, 303.
24. Hadjidemetriou J.D. (1975): *The continuation of periodic orbits from the restricted to the general three-body problem*, Cel. Mech. **12**, 155-174.
24. Hadjidemetriou J.D. (1976): *Families of Periodic Planetary Type Orbits in the Three-Body Problem and their Stability*, Astrophys. Space Science **40**, 201.-224.
24. Hadjidemetriou J.D. (2002): *Resonant periodic motion and the stability of extrasolar planetary systems*, Cel. Mech. & Dyn. Astron. **83**, 141-154.
24. Israelinian G., Santos N, Mayor M. and Rebolo R. (2001): *Evidence for planet engulfment by the star HD82943*, Nature **411**, 163.
24. Ji, J., Kinoshita, H., Liu, L., Li, G. and Nakai, H. (2003): *The apsidal antialignment of the HD 82943 system*, Proceedings of IAU Col. No 189, (accepted to Cel. Mech. & Dyn. Astron).
24. Jones B. and Sleep P. (2002a): *The stability of the orbits of Earth-mass planets in the habitable zone of 47 Ursae Majoris*, Astron. Astrophys. **393**, 1015.
24. Kinoshita H. and Nakai H (2001a): *Stability of the GJ 876 Planetary System*, PASJ,**53**,L.25-L.26.
24. Kinoshita H. and Nakai H.(2001b): *Stability Mechanism of Planetary System of $v'$ Andromedae*, Proceedings of the IAU Symposium 202, Manchester 2000.

24. Kiseleva-Eggleton L., Bois E., Rambaux N. and Dvorak R. (2002): *Global dynamics and stability limits for planetary systems around HD 12661, HD 38529, HD 37124 and HD 160691*, ApJ. Letters, 578, L145.
24. Krzysztof Gozdziewski, Eric Bois and Andrej J. Maciejewski (2002): *Glaobal dynamics of the Fgliese 876 planetary system*, Mon. Not. R. Astron. Soc.
24. Laughlin G. and Chambers J. (2001): *Short-term dynamical interactions among extrasolar planets*, Ap.J. **551**, L109-113.
24. Lissauer J.J. and Rivera E.J.(2001): *Stability analysis of the Planetary System orbiting Andromedae*, Ap.J. **554**, 1141-1150.
24. Malhotra R. (2002a): *A dynamical mechanism for establishing apsidal resonance*, ApJ. Letters bf 575.
24. Malhotra R. (2002b): *Eccentricity excitation and apsidal alignment in exoplanetary systems*, in 34th DPS Meeting, BAAS, 34, 42.05
24. Murray N., Paskowitz M. and Holman, M. (2001):*Eccentricity evolution of resonant migrating planets*, Ap.J. (preprint).
24. Peale S. and Lee M. (2002): *Extrasolar Planets and the 2:1 Orbital Resonances*, in DDA 33rd Meeting, BAAS, 34, #1.02
24. Rivera E.J. and Lissauer J.J. (2001): *Dynamical models of the resonant pair of planets orbiting the star GJ 876*, Ap.J. **558**, 392- 402.
24. Schneider Jean, (2002): *http://www.obspm.fr/encycl/catalog.html*.

# The "Third" Integral in the Restricted Three-Body Problem Revisited

Harry Varvoglis, Kleomenis Tsiganis, and John D. Hadjidemetriou

Department of Physics, University of Thessaloniki, GR-541 24 Thessaloniki, Greece

**Abstract.** In 1964 M. Hénon and, independently, V. Szebehely with G. Bozis presented the first numerical results, indicating the existence of a "new" local integral of motion in the circular restricted three-body problem. The first terms of an asymptotic expansion of this integral were later calculated by Contopoulos [1]. Several years later, the Celestial Mechanics astronomical community started to develop a very successful theory on local integrals of motion in the restricted three-body problem, which however in the jargon of this field are called *proper elements* and are related to known analytical approximate solutions. The calculation of proper elements is based on the implicit assumption that the orbit traced by a planet (major or minor) is nearly-regular. Here we show that this method is also applicable, albeit partly, in a special case of chaotic motion in the Solar System, known as "stable chaos". Thus, the existence of an additional local integral of motion in the elliptic restricted three-body problem is responsible for the phenomenon of stable chaos.

## 1 Introduction

In 1964 the Laboratory of Astronomy of the University of Thessaloniki hosted IAU Symposium 25. This meeting was devoted to the interaction between astronomers working on two widely different fields of Dynamical Astronomy, namely Galactic Dynamics and Celestial Mechanics, in the hope that the methods used traditionally in one of the fields could prove useful in the other. Indeed, several papers presented in this meeting followed the above line. In two of them Hénon [2], on the one hand and, independently, Szebehely and Bozis [3] on the other, reported that they had found indications for the existence of a further integral of motion in the planar circular restricted three-body problem (a two-degrees of freedom dynamical system), besides the well known Jacobi integral.

Subsequently Contopoulos [1] showed how this integral could be constructed in a series form through an algorithm similar to the one he had proposed already [4] for the "third" integral in the case of a galactic type potential, in which (series) the zeroth order term is the angular momentum. At the same time Bozis [5] [6] studied extensively the properties of this new integral, as well as the computation, through its use, of "generalized" elements of motion (e.g. eccentricity, see next paragraphs).

Since Poincaré had shown that the three-body problem is non-integrable, it is obvious that this integral can only be a "local" (non-isolating) one. Therefore one should inquire in which regions of phase space this integral may be applied, as it was initiated by Bozis [6]. These regions should be called "regular", since the

G. Contopoulos and N. Voglis (Eds.): LNP 626, pp. 433–441, 2003.

corresponding dynamical system has two degrees of freedom and, therefore, in the regions where there exist two integrals of motion, it behaves like an integrable one.

A three-dimensional elliptic orbit of the two-body problem is uniquely defined by three quantities, the three *elements* of the orbit $a$, $e$ and $I$, where $I$ is the *inclination* of the plane of the orbit with respect to a "reference" plane, $a$ is the semi-major axis of the ellipse and $e$ the *eccentricity.* In what follows we consider the motion of massless test-particles (i.e. asteroids) relative to a massive central body (i.e. the Sun) of mass $M$. The orbital elements of the minor planet are related to the energy, $E$ and the angular momentum, $h$, of its orbit, through the relations

$$a = -\frac{G\,M}{2E} \tag{1}$$

$$e = \sqrt{1 + \frac{2Eh^2}{G^2 M^2}} \tag{2}$$

For elliptic motion, the orbital energy, $E$, has to be negative.

It is worth to note that the two-body problem is an intrinsically degenerate dynamical system [9], a property that becomes obvious if we write the corresponding Hamiltonian in action-angle variables. One possible set of action angle variables in this case are the well known *modified Delaunay variables*, defined through the relations

$$\Lambda = \sqrt{G\,M\,a} \quad \lambda = \varpi + l \tag{3}$$

$$\Gamma = \Lambda(1 - \sqrt{1-e^2}) \quad \gamma = -\varpi \tag{4}$$

$$Z = \Gamma(1 - \cos i) \quad \zeta = -\Omega \tag{5}$$

where the angles $\Omega$, $\varpi$ and $l$ are the three Euler angles: the first two define the orientation of the ellipse in space and the third one the position of the the asteroid on the ellipse. In Celestial Mechanics the various angles have their own names: $\Omega$ is the *longitude of the ascending node* of the orbit, $\varpi = \Omega + \omega$ is the *longitude of the pericenter* and $\lambda = \varpi + l$ is the *mean longitude.* The *mean anomaly*, $l$, is related to time through the relation $l = n\,t$, where $n$ is the *mean motion of the planet*, i.e. its mean angular frequency around the massive central body. The Hamiltonian of the two-body problem, written in the above variables, becomes simply

$$H = -\frac{G^2\,M^2}{2\,\Lambda^2} \tag{6}$$

i.e. it depends only on the action corresponding to the energy, which, according to (1), depends only on the semi-major axis.

The two-body problem is only a simple approximation of a planet's motion around the Sun. A better approximation is the restricted three-body problem. In this model a massless particle is moving in the gravitational field of two bodies, a central massive primary of mass $M$ (the Sun) and a perturbing planet of mass $m$ (say Jupiter). Moreover, the motion of the perturbing planet around the Sun is a Keplerian closed orbit (i.e. either a circle or an ellipse). The trajectory of the massless body is not anymore an ellipse, due to the perturbations induced by the planet. However, due to the small mass of the perturber relative to the Sun and for relatively large separation between the asteroid and the perturber, the trajectory can be described by means of the *osculating elements*, i.e. instantaneous values of the variables $a(t)$, $e(t)$ and $I(t)$, defined as the elements of an ellipse that is tangent to the real orbit at time $t$. The process is very easily implemented, since it reduces to the calculation of the elements of the orbit from the instantaneous values of the energy and the angular momentum (which, of course, are not anymore constants in the case of the three-body problem).

## 2 The Way Things Might Have Happened

### 2.1 Ordered Trajectories

From the form of the Hamiltonian alone and some educated guesses, one could relatively easily arrive at the form of the third integral, for the existence of which Hénon, Szebehely and Bozis had found numerical evidence, as follows. In the restricted three-body problem the Hamiltonian can be "split" into two parts, one of order zero with respect to the *mass ratio*, $\mu = \frac{m}{M+m}$, and one of order unity. In modified Delaunay variables the zeroth-order term depends only on $\Lambda$, while the other two actions appear only in the first order term, which therefore may be considered as a "perturbation". Thus, we have again a case of degeneracy, similar to the one appearing in the two-body problem. Due to this degeneracy, the Fourier expansion of the perturbation contains terms that do not depend on the angle $\lambda$. Therefore, if one ignores the terms involving $\lambda$ and $\lambda'$ [1], which become important only when they are almost resonant, the osculating semi-major axis, $a$, is constant, a famous result known as the *Laplace-Lagrange linear theory of secular motion.* Then $E$ is constant to a linear approximation as well, since it depends only on the osculating semi-major axis through (1). As a consequence and, in view of (2), the osculating eccentricity, $e$, is, to a linear approximation, a function of $h$ only, i.e. $e$ depends, essentially, only on the angular momentum. Therefore it is natural to expect that, if one would attempt to calculate a "third" integral *for the full, non-linearized problem* as a series, using as a small parameter the mass ratio, $\mu$, the zero-order term should be the angular momentum of the massless body on its (unperturbed) orbit around the central body. This is exactly the method used by Contopoulos [1]. In the same linear approximation as for $a$, the osculating eccentricity of the asteroid is given by

$$e^2 = e_f^2 + e_P^2 + 2\, e_f\, e_P\, \cos(g_P\, t + \beta_P), \tag{7}$$

where $e_f$, $e_P$, $g_P$ and $\beta_P$ (the phase at $t = 0$) are constants. In particular $e_f$ (usually called *forced eccentricity*) and $g_P$ (*proper frequency*) depend only on $a$ and $\mu$, while $e_P$ is the constant amplitude of variation of the osculating

[1] Note that by a prime we denote the angles of the perturbing planet

eccentricity. In the full, non-linearized problem, $e_P$ can be calculated through an algorithm similar to the one used by Contopoulos [1], and is called the *proper eccentricity.*

Since the circular restricted three-body problem is a two-degrees of freedom autonomous dynamical system, the existence of a second integral of motion would imply integrability. In this case all trajectories would be ordered and the secular solution would always remain $\mathcal{O}(\mu)$ close to the real solution. Note that (7) is the simplest secular theory of Celestial Mechanics (e.g. see Yuasa [7] or Milani and Knežević [8]). This result can be generalized for forms of the restricted three-body problem with more than two degrees of freedom, such as the elliptic (where the orbit of the perturber is an ellipse) or the three-dimensional (where the massless body moves outside the plane of the orbit of the perturber). In these cases one would need to calculate further integrals of motion, in the same spirit. As far as the total number of integrals is equal to the number of degrees of freedom of the corresponding (autonomous) dynamical system, all trajectories would be ordered. In this way we see that the three proper elements of the trajectory (or the associated modified Delaunay variables) constitute a set of action variables (and hence integrals of motion) of the *secular three-body problem.*

## 2.2 Chaotic Trajectories

The proper elements of ordered trajectories of asteroids are calculated through the secular theory at any desirable level of accuracy. However we know, from the work of Poincaré, that the restricted three-body problem does not admit any further integrals of motion, *analytic* in any variables. Therefore the corresponding dynamical system is non-integrable and the integrals in series form calculated through the method of Contopoulos (or some secular theory) can only be non-isolating, local ones. Hence in the vicinity of orbital resonances between the test-particle and the perturber (i.e. resonances between the angles $\lambda$ and $\lambda'$) the secular theory should fail, as a result of the small divisors problem and the appearance of chaotic motion. This means that all specific models of the restricted three-body problem (e.g. circular, elliptic or three-dimensional) should possess chaotic phase-space regions, besides the ordered ones. What can we say on the properties of chaotic trajectories? This problem was attacked by many authors through extensive numerical calculations, according to the available, at any period, computing power. The first model studied was the simplest one, namely the planar circular restricted three-body problem.

Soon it was realized, however, that this model does not represent the generic case, since it corresponds to an autonomous dynamical system with two degrees of freedom. But in this class of dynamical systems Arnold's diffusion (see e.g. [9]), which might play an important role in solar system dynamics, cannot be taken into account. Therefore, if we would like to consider a "generic" model for three-body dynamics, we should have at least three degrees of freedom! Consequently one should use as a "generic model" either the elliptic planar restricted or the circular three-dimensional restricted problem and not the planar circular. This was done by Contopoulos, who calculated the form of the "third" integral in the case of the three-dimensional restricted three-body problem [10] and the planar elliptical three-body problem [11].

The difference between the circular restricted three-body problem, on one hand, and the three-dimensional or elliptic restricted problem, on the other, is qualitative[2]. In both cases there exists a global (isolating) integral, which is the Jacobi integral in the first and the Hamiltonian of the extended phase-space in the second. But in the first case the situation is clear-cut: a specific trajectory is either ordered (if an additional local integral exists) or chaotic (if no local integrals exist). In the second case, however, there may exist from none to two local integrals of motion [12]. Two local integrals imply regular behavior and ordered trajectories, for which the secular solution would be an accurate approximation. The other two sub-cases correspond to chaotic motion, but with significant differences. If no local integrals exist, the chaotic trajectory covers densely a sub-manifold of the phase-space, defined by the constant "energy" surface. If one local integral exists, then the trajectory lies on a manifold which is the cartesian product of a two-dimensional torus with an annulus [18] (see Fig. 1). The motion on the two-torus corresponds to the ordered part of the trajectory, originating from the existence of the two integrals, while the motion on the annulus corresponds to the chaotic part.

In the case where no local integrals exist, the motion is "fully" chaotic, i.e. macroscopically it is equivalent to a random walk. Therefore, one might use methods of statistical mechanics (e.g. a Fokker-Planck-type equation) in order

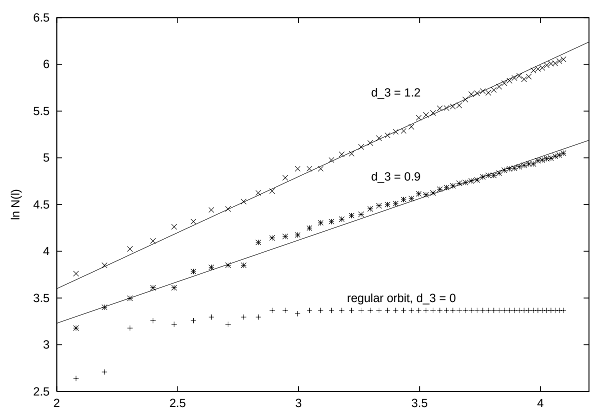



**Fig. 1.** Calculation of the number of integrals of three trajectories, one ordered and two stable chaotic, in the region of the 12:7 orbital resonance (from [15]). According to the theory, if we partition a 3-$D$ space in $M^3$ bins of side $l$, $N$ of which are occupied by a trajectory, then we have that $\log N(l) \sim d\_3 \log M(l)$, where $d\_3 = 3 - d$, and $d$ is the number of integrals. The regular orbit yields $d\_3 = 0$, i.e. $d = 3$, while stable-chaotic orbits have $d\_3 \approx 1$, i.e. $d \approx 2$

[2] The 2-D elliptic and the 3-D circular problem are also by no means equivalent to each other.

to describe the evolution of a set of initial conditions as a diffusion process in the elements space. Since, according to what has been already said, the semi-major axis is constant to a linear approximation, we can select as a dependent variable either the eccentricity or the inclination. The eccentricity is our first choice, since it is intimately related to the escape of asteroids from the main belt.

It is easy to see that $e$ increases on the average, since if we consider the chaotic motion as a random walk in eccentricity space, there is a reflecting wall at $e = 0$! Moreover, as $e$ increases the resonances begin to overlap and chaotic motion becomes dominant. Therefore asteroids in fully chaotic trajectories follow more and more elongated orbits, until they hit a planet and are removed from the distribution. An analytic theory for the diffusion of asteroids was developed by Murray and Holman [13] and was recently applied, with considerable success, for the estimation of the age of the Veritas family of asteroids [14].

In the case where one local integral exists, the motion is "partially" chaotic, which means that some degrees of freedom are evidently chaotic and some appear as being ordered. From extensive numerical experiments it is relatively straightforward to show that the evolution of $a$ is chaotic, while $e$ and $I$ change almost quasi-periodically with time, their proper values being almost constant [16] [17] [18] (Fig. 2). But, according to the secular theory, $a$ only undergoes bounded erratic oscillations and does not change secularly, unless of course the trajectory escapes from the (non-isolated) region of the elements' space, where it is restricted by the level surfaces of the local integral. Since the usual way for the classification of trajectories is through the calculation of the Maximal Lyapunov Number, which in this case is positive, "partially chaotic" trajectories could be named, as well, "stable chaotic". Since for a stable chaotic trajectory $e_P$ does not increase on the average, there are no collisions with other planets and, therefore, no escapes.

Extensive numerical work has shown that another important property of a phase-space region, besides the existence of local integrals of motion, is the existence or not of simple-periodic resonant trajectories. Although in the restricted circular three-body problem all orbital resonances with Jupiter correspond to periodic trajectories, this is not true for the elliptic problem. In general, orbital resonances do not correspond to periodic trajectories, unless their period is an exact multiple of Jupiter's revolution period [16]. Thus, the chaotic regions of phase space (i.e. the resonances' zones), in the planar elliptic (or the three-dimensional circular) restricted three-body problem, can be classified into three classes as follows, according to the type of trajectories they contain and the existence or not of periodic trajectories [16] [17] [18].

Stable chaotic regions constitute the first class. In such a region the evolution of trajectories is not diffusive. Chaotic trajectories are semi-confined by the level surfaces of the local integrals. Since, however, these surfaces are non-isolating, the trajectory eventually escapes from such a region through the "holes" of the "invariant" manifold. After such an escape, the eccentricity increases steeply. Numerical experiments have shown that the typical time-scale, $T$, for escape

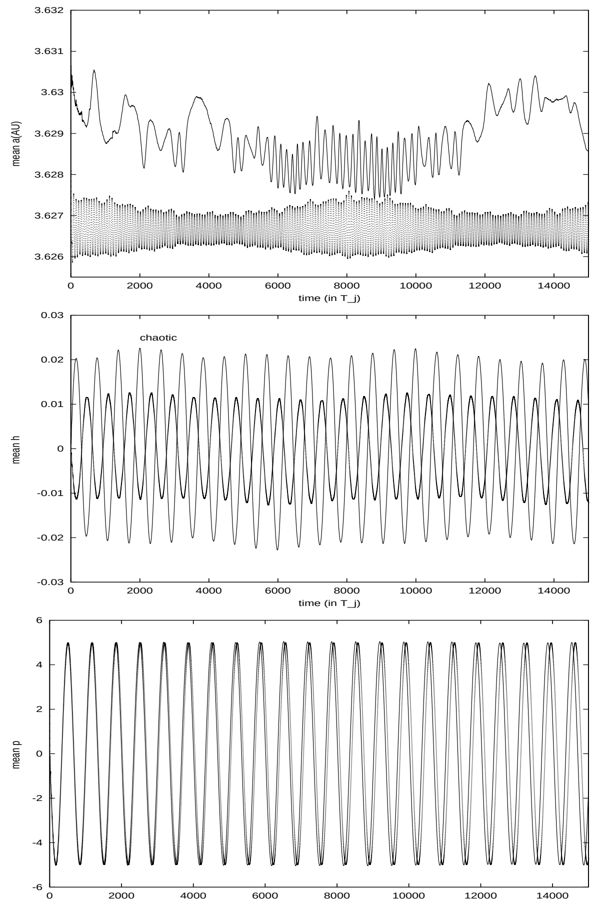



**Fig. 2.** The elements $a$ (top), $h = e \sin \varpi$ (middle) and $p = I \sin \Omega$ (bottom) are given, as functions of time, for one regular and one stable-chaotic orbit of the elliptic three-body problem in the vicinity of the 12:7 orbital resonance (from [15]). The unit of time is the revolution period of Jupiter, $T_J \approx 11.86$ yr. It is easy to realize the different character of the motion between these two orbits, by monitoring the behavior of $a$. On the other hand, one cannot decide whether an orbit is regular or chaotic by just observing the graphs of $h$ or $p$

through this process is $T \sim 1$ Gyr and can even exceed the age of the solar system (5 Gyrs), depending on the specific resonance.

Fully chaotic regions are divided into two classes, according to whether they support simple periodic orbits or not. If there are no periodic orbits, the evolution is diffusive, i.e. a trajectory undergoes many small "jumps" in eccentricity. This case is the one that can be described successfully through a diffusion equation and its typical time-scale, as can be calculated by the values of the diffusion coefficient, is of the order of 100 Myrs $< T <$ 1,000 Myrs (again, depending on the specific orbital resonance).

If there exist periodic orbits, then the evolution of chaotic trajectories is "fast" and intermittent, as the trajectory from time to time follows the unstable periodic orbit. This is the kind of motion found by Wisdom [19] and Hadjidemetriou [20]. The typical time-scale for the "jumps" is of the order of $5 \cdot 10^5$ yrs, while the escape time is of order $10^5 < T < 10^6$ yrs. There are only 5 such resonances in the phase-space region that corresponds to the main asteroid belt, in both the elliptic and the three-dimensional restricted three-body problems. These are the 2:1, 3:1, 4:1, 5:2 and 7:3 orbital resonances with Jupiter. Since the more well-known Kirkwood gaps lie exactly at these resonances, one arrives easily at the conclusion that the existence of a periodic trajectory is the common factor that differentiates between orbital resonances, associated with a Kirkwood gap, and those that are not.

Summarizing, we can say that stable chaos is the observational manifestation of the existence of a local integral of motion, while the Kirkwood gaps appear at resonances where periodic orbits exist, in the elliptic or the three-dimensional restricted three-body problem.

## 3 The Way Things Really Happened

Unfortunately, the evolution of ideas in science does not always follow the "obvious" path. The applicability of local integrals of motion presents another case of misunderstanding between theorists and applied-oriented astronomers. The scientific community of Celestial Mechanics did not capitalize on the work of Bozis and Contopoulos, related to the existence of local integrals of motion and the calculation of "primitive proper elements". Instead, for quite some time, the calculation of proper elements was only used for objects that move far away from the main resonances, where secular theory could apply.

Things started to change in the 1980's, when algorithms for the calculation of the maximal LCN were made available and Wisdom [19] found the "intermittent" behavior of the osculating eccentricity in the vicinity of the 3:1 resonance, which is characteristic of the existence of an unstable periodic trajectory. However, since as a rule only the maximal LCN was calculated, there was no way to differentiate between regions where none or one local integral exists. That is why the chaotic motion in the regions where local integrals exists was considered "peculiar" and termed *stable chaos*.

The first to point out that stable chaotic motion is not "fully chaotic" were Varvoglis and Anastasiadis [21]. This idea was subsequently elaborated in a series of papers by Tsiganis, Varvoglis and Hadjidemetriou [16] [17] [18]. In these papers it is shown, through the computation of autocorrelation functions, that stable-chaotic trajectories have almost constant proper elements, i.e. they possess local integrals of motion (see Fig. 1), and lie at the border between fully chaotic and regular phase-space regions. Consequently, stable-chaotic orbits represent cases of *sticky motion* in $G$ and $H$ (i.e. essentially eccentricity and inclination) and chaotic motion in $L$ (i.e. semi-major axis), a type of motion for which no analogue exists in two-dimensional dynamical systems. The subsequent numerical calculation of the number of integrals, preserved by a large number of trajectories of the elliptic restricted three-body problem [15], confirmed this picture. In this way today we arrived finally, after thirty-six years, in the “re-discovery” of the work of Contopoulos-Bozis and its connection to proper elements, by understanding the phenomenon of stable chaos and its relation to local integrals of motion.

## References


1. G. Contopoulos: Astrophys. J. **142**, 802 (1965)
2. M. Hénon: “Numerical exploration of the restricted three-body problem”. In: *The theory of orbits in the Solar System and in stellar systems, IAU Symposium No. 25, Thesaloniki, Greece, 17–22 August 1964*, ed. by G. Contopoulos (Academic Press, London 1966) pp. 157–163
3. V. Szebehely: Discussion of the paper “Numerical exploration of the restricted three-body problem”. In: *The theory of orbits in the Solar System and in stellar systems, IAU Symposium No. 25, Thesaloniki, Greece, 17–22 August 1964*, ed. by G. Contopoulos (Academic Press, London 1966) pp. 163–169
4. G. Contopoulos: Z. Astrophys. **49**, 273 (1960)
5. G. Bozis: Astron. J. **71**, 404 (1966)
6. G. Bozis: Astron. J. **72**, 380 (1967)
7. M. Yasa: Publ. Astron. Soc. Japan **25**, 399 (1973)
8. A. Milani, Z. Knežević: Cel. Mech. **49**, 247 (1990)
9. A.J. Lichtenberg, M.A. Lieberman: *Regular and Chaotic Dynamics*, 2nd edn. (Springer, New York 1992)
10. G. Contopoulos: Astron. J. **72**, 191 (1967)
11. G. Contopoulos: Astron. J. **72**, 669 (1967)
12. G. Contopoulos, L. Galgani, A. Giorgilli: Phys. Rev. A **18**, 1183 (1978)
13. N. Murray, M. Holman: Astron. J. **114**, 1246 (1997)
14. Z. Knežević, K. Tsiganis, H. Varvoglis: “The dynamical portrait of the Veritas family region”. In: *Proceedings of the Conference Asteroids, Comets, Meteors - ACM2002, Technical University Berlin, 29 July-2 August 2002*, ed. by B. Warmbein (ESA SP-500, Noordwijk, 2002) pp. 335–338
15. H. Varvoglis, K. Tsiganis, G. Hadjivantsides: “Stable chaos and local integrals of motion”. In: *Proceedings of the Conference Asteroids, Comets, Meteors - ACM2002, Technical University Berlin, 29 July-2 August 2002*, ed. by B. Warmbein (ESA SP-500, Noordwijk, 2002) pp. 355–357
16. K. Tsiganis, H. Varvoglis, J. Hadjidemetriou: Icarus **146**, 240 (2000)
17. K. Tsiganis, H. Varvoglis, J. Hadjidemetriou: Icarus **155**, 454 (2002)
18. K. Tsiganis, H. Varvoglis, J. Hadjidemetriou: Icarus **159**, 284 (2002)
19. J. Wisdom: Astron. J. **87**, 577 (1982)
20. J. Hadjidemetriou: Celest. Mech. Dyn. Astron. **56**, 563 (1993)
21. H. Varvoglis, A. Anastasiadis: Astron. J. **71**, 404 (1996)